DOCUMENT RESUME

ED 236 266 UD 023 111

AUTHOR Greenberg, Karen
TITLE The Effects of Variations in Essay Questions on the Writing Performance of CUNY Freshmen. Research Monograph Series No. 1.
INSTITUTION City Univ. of New York, N.Y. Office of Academic Affairs.
PUB DATE Jan 81
NOTE 103p.; For related documents, see UD 023 109 and UD 023 112
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Diagnostic Tests; Educational Diagnosis; Evaluation Criteria; Evaluation Methods; Higher Education; Holistic Evaluation; *Minimum Competency Testing; Test Reliability; *Test Validity; Undergraduate Students; *Writing (Composition); *Writing Evaluation
IDENTIFIERS City University of New York; *Writing Skills Assessment Test

ABSTRACT

The reliability of the Writing Skills Assessment Test (WSAT) given to all entering freshmen at the City University of New York (CUNY) is the subject of this monograph. The monograph states that although the WSAT is scored according to validated and widely-used holistic procedures, some CUNY faculty had expressed concern about aspects of the WSAT's question wording. The study reported on here experimentally manipulated two aspects of the WSAT questions in an attempt to determine whether variations would increase the test's potential for eliciting students' minimal writing competencies. The aspects manipulated were (1) cognitive demand levels--the degree of information and the number of strategies which an essay question provides; and (2) experiential demand levels--the degree of personal experience for which an essay question asks. The results showed that neither of the changes had any significant effects on students' writing performance; these findings provide empirical construct validation of the WSAT. Appendices to this monograph contain a sample WSAT form, the evaluation scale used with the WSAT, sample (annotated) student essays representing each point on the scale, the essay questions used in this study, and inter-coder reliability data. (CMG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marie Jean Leterman

CUNY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- ✓ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CONTENTS

Introduction 1

Measurement and Operational Definitions 6

Methodology 28

Results 36

Discussion and Conclusions 58

Pedagogical Implications 65

References 67

Appendices

A: Sample FSAP essay form 76

Evaluation Scale 77

Annotated student essays 78

B: Essay questions used in study 94

C: Inter-coder reliability data..... 98

INTRODUCTION

In the fall of 1978, the City University of New York mandated an essay test to be given to all entering freshmen for the purpose of accurate diagnosis of writing ability and proper placement in writing courses. Students who do not meet the minimum standard established by the university must take and pass an alternate form of the test after remediation and before they enter the junior year. A sample of the test and the evaluation scale constructed by CUNY basic skills faculty, and used by the faculty raters who score the essays, can be found in Appendix A. Basically, each essay is read by two readers and rated holistically on a scale of 1 through 6. Sample student essays representing each point on the scale are also included in Appendix A. A student's final score then becomes the summed score of the two individual ratings, with a possible range of 2 through 12. Students must achieve a total score of at least 8 in order to be certified as meeting the University's minimum standard.

This study attempted to contribute to the relatively unexplored area of essay test question validation and to respond to the concerns expressed by some writing faculty about the CUNY Writing Skills Assessment Test (WSAT). The study focused on the writing of students at one of the eight senior colleges within the seventeen undergraduate colleges of the University system.

At this college, the WSAT has proven to be a very effective means of ensuring that students entering the college-level freshman composition courses have the writing competencies necessary to function in these courses. Moreover, the faculty is aware of the WSAT's validity and reliability and of the fairness which these two psychometric properties provide for the evaluation of students' writing.

The WSAT has high construct validity because it consists entirely of a writing sample, and, thus, it requires students to demonstrate the skills and abilities associated with the construct being assessed: minimal writing competencies for college-level work. In addition, the test's designers were very clear about what constitutes minimal writing competence: "the lowest level of writing ability which students should have in order to take advantage of what they will be taught in college-level composition" (CUNY Task Force on Writing, 1979, p. 1). The specific skills and processes which a writer at this level should possess are clearly and comprehensively articulated in the 4-5 section of the CUNY Evaluation Scale.

In terms of the reliability of the WSAT, the test is scored according to the holistic procedures created and validated by measurement specialists and used by many universities, colleges, and testing agencies (Cooper, 1977). However, some faculty have expressed concern about the appropriateness

of selected aspects of the WSAT's question wording. Many teachers object to the unvarying "agree/disagree" format and to the impersonal nature of some of the question topics. These objections should be examined in a series of carefully controlled experimental studies. While it is reasonable to wonder about aspects of a test's validity, it is not reasonable, in terms of psychometrics, logic, economy, or practicality, to alter any aspect of a test without providing statistically significant evidence for that change.

Thus, this study was undertaken to determine whether changes in the wording of the WSAT questions would elicit superior rhetorical, syntactic, and lexical performance from a sample of senior college freshmen.

More specifically, the purpose of this study was to determine the effects of variations in two aspects of the wording of the WSAT questions on students' writing performance. The two aspects of the questions which were experimentally manipulated were the questions' "cognitive" and "experiential" demand levels. "Cognitive demand level" refers to the degree of information and the number of strategies which an essay question provides. "Experiential demand level" refers to the degree of personal experience for which an essay question asks. Both of these independent variables will be operationally defined in the next section.

The categories of writing performance (the dependent variables) which were examined in this study were overall writing quality, syntactic complexity, sentence control errors, vocabulary errors, and essay length. They, too, will be operationally defined in the next section.

A secondary purpose of this study was to determine the relationships between the essays' holistic quality scores and the essays' syntactic complexity, sentence control errors, vocabulary errors, and essay length. The examination of these relationships may provide writing teachers with information about the specific cues in students' writing which affect readers' responses.

In order to examine all of the above relationships, the following hypotheses were proposed:

Hypothesis One

There will be a significant interaction between the cognitive and the experiential demand levels of the essay questions with Low Cognitive/High Experiential questions eliciting higher overall writing quality.

Hypothesis Two

There will be a significant interaction between the cognitive and the experiential demand levels of the essay questions with Low Cognitive/High Experiential questions eliciting greater syntactic complexity, fewer sentence control errors, fewer word form errors, and longer essays.

Hypothesis Three

There will be a significant difference in the syntactic complexity, the number of sentence control errors, the number of word form errors, and the number of words between essays assigned failing holistic quality scores (2 through 7) and essays assigned passing holistic quality scores (8 through 12).

Hypothesis Four

The essays' assigned holistic quality scores will be significantly related to the quantitatively measured characteristics (of syntactic complexity, number of sentence control errors, number of word form errors, and number of words).

MEASUREMENT AND OPERATIONAL DEFINITIONS

The question format of the CUNY Writing Skills Assessment Test includes the following:

1. A short passage (20-60 words) sometimes consisting of a direct quotation.
2. The question, "Do you agree or disagree?"
3. The direction, "Explain and illustrate your answer from your own experience, your observations of others, or your reading."

Sample essay questions appear in Appendix A. The subject matter of the questions on various forms of the test has included advertising, gun control, housing, job security, juvenile crime, marriage, New York City, school, sex roles, spectator sports, success, television, women and the draft, and writing. Students are asked to choose one of two topics and write a fifty-minute essay. The basic requirement of the test is that students take a position on the topic and defend that position with a persuasive "argument."

There is evidence provided by a body of research on the effects of different rhetorical modes showing that argumentative writing calls for more syntactic complexity than do other modes. Studies conducted by Britton, Burgess, Martin, McLeod, and Rosen (1975), Crowhurst and Piché (1979), Perron (1976a, 1976b, 1977) and San Jose (1972) showed that

students produce more syntactically complex writing when they respond to questions eliciting arguments than when they respond to questions eliciting explanations, descriptions, or narrations. There are many other question "cues," however, which can be used to elicit argumentative responses besides "agree or disagree." Cue phrases such as "discuss," analyze," "explain," and "show why," can also be used to stimulate students to assert and defend positions in unified, logical essays.

This study attempted to determine if, in fact, the "agree/disagree" format does have significant effects on students' writing performance. The study also attempted to control several potential sources of bias in the wording of the questions in the following ways:

1. Each passage was submitted to a panel of four composition teachers who agreed that each passage made only one assertion about the topic.
2. Each passage had approximately the same level of syntactic difficulty according to the Syntactic Density Score (Golub, 1973) with Belanger's modifications for balanced weightings (Belanger, 1978). The scores on this measure ranged form 3.5 to 4.1.

Thus, an attempt was made to control the rhetorical, propositional, and syntactic difficulty of the experimental questions in order to isolate and determine the effects of the two variables under study: the cognitive demands and the experiential demands of the questions.

Cognitive Demands

For this study, "High Cognitive Demands" questions were ones which could be classified as "evaluative," and "Low Cognitive Demands" questions were ones which could be classified as "interpretive." According to Sanders (1966), "evaluative" and "interpretive" questions may be similar in that both types require "the ability to see a relationship between an inductive conclusion and supporting evidence" and in both types, "the student is given an inductive conclusion and asked to find or recognize evidence that supports it" (Sanders, 1966, p. 52). The major difference between the two types, however, concerns the question's specification of criteria for the evidence.

Evaluative questions require students to perform two cognitive steps: (1) set up the appropriate criteria or standards and (2) determine how closely the idea or object meets these standards (pp. 141-142). They require conscious value judgements based on students' own clearly defined criteria (Gronlund), 1970, p. 20). Interpretive questions, on the other hand, specify the criteria to be used in making a judgement (Sanders, 1966, p. 142). They provide students with generalizations and with the criteria for finding evidence that relates to the truth or falsity of the generalization (p. 52) or they ask students to discuss the relationships between the generalizations, issues, or objects (p. 43).

They require the ability to relate what is given in the question to past, present, and future conditions (Gronlund, 1970, p. 20).

According to Sanders, evaluative questions make more difficult cognitive demands on students than do interpretive questions because students need more sophisticated intellectual abilities to answer them. More specifically, in order to "evaluate" an issue, students must be able to distinguish between "facts," "opinions," and "values" and be able to discuss all three. Also, they have to be able to evaluate the issue from all sides even if it concerns values which are strongly influenced by their cultural, religious, or ethnic biases. And finally, they must be able to see beyond their own individual values and know that other people may cherish very different values (pp. 143-145).

It is important to note that the concept of "evaluation," as Bloom (1956) and Sanders (1966) refer to it, involves "an adequate comprehension and analysis of the phenomena to be appraised" and "a fully conscious awareness of the criteria for forming these appraisals" (Bloom, 1956, p. 186). Bloom distinguishes this type of an appraisal from "opinions" which are "quick decisions not preceded by very careful consideration of the various aspects of the object, idea, or activity being judged" (p. 186).

The questions on the writing test which were used for this study ask for "evaluations" as opposed to "opinions" because they require students to "agree or disagree" and to defend their position with logical reasons. In order to do this, students have to take one position on the issue and articulate the criteria on which they are basing their agreement or disagreement. This may be a formidable task to accomplish in a 50-minute essay test, and it is likely that under the time pressure, many students may respond with "opinions" rather than with carefully considered "evaluations."

An example of a question which asks for a complex "evaluation," but which probably evokes a highly emotional "opinion," is the following:

> In this country, the educational process takes far too long. We have all gotten used to compulsory attendance laws and to twelve years of precollege schooling, but most students could learn everything schools have to teach them in about half the time. Do you agree or disagree?

This question (an actual exam question) is extremely complex. In order to give an answer which will receive a high score, students may have to consider why the educational process takes twelve years, whether it should be shortened and why or why not, and what kinds of changes this "shortening" would cause (for example, the type of changes in

schools, classes, and teachers which would have to be effected and the problem of the millions of teenagers who would be entering the job market at an earlier age).

An interpretive question, on the other hand, does not pose so many complex cognitive demands:

> A characteristic of interpretation is that the student discovers or uses relationships on a common-sense level. The emphasis in the question is on finding the relationship among the parts of the subject matter and not on formal understanding of the thought processes involved.
>
> (Sanders, 1966, p. 43)

Sanders cites several different types of interpretive questions: questions asking for comparisons and contrasts, deductions and inductions, definitions and classifications, and causes and effects. Thus, an example of an interpretive version of the evaluative question discussed above is as follows:

> In this country, there are compulsory attendance laws which require students to attend school from first grade through twelfth. This is a much longer time than the six to nine years of schooling required by many other countries of the world. In an essay of about 300 words, discuss our compulsory educational process. In your essay, you might wish to choose one of the following strategies: contrast our educational system to that of another country, or explain some reasons for and effects of twelve years of compulsory school attendance, or suggest changes in the present attendance laws and give reasons for these changes.

This type of question requires students to relate ideas rather than judge them. Furthermore, it asks students to "discuss" their ideas instead of requiring them to "agree or disagree." And most importantly, it offers students a variety

of equally legitimate intellectual options for structuring an answer. Thus, each student who answers an interpretive question can select the option with which he or she is most comfortable (in intellectual terms), and he or she is not locked into agreeing or disagreeing with a simplistic judgement on a complex issue.

Experiential Demands

Because essay tests of writing are not aimed at measuring students' knowledge or mastery of specific content, the subject matter of these questions is usually general, presenting ideas of objects with which most respondents are assumed to be familiar. However, some essay tests of writing ask questions about subject matter which may be totally outside the range of some students' experiences or concerns. If a student does not have the information necessary to answer the question, the question may interfere with his or her ability to produce minimal writing competencies.

Little research on the effects of different types of subject matter on students' writing performance exists. This is extremely surprising considering how important this variable may be. As Emig (1971) noted in her descriptive study of student writers,

> if a teacher gives an assignment requiring writing in this sub-mode (on an abstract issue), certain students may be unable to complete adequately, or even begin, such an assignment. Along with being intellectually unable to perform the assignment, the student may also be unmotivated or psychically unable to perform the assignment. Such "blocks"

> emanate from strikingly different sources: the student may find the task too boring or he may find the task too threatening (p. 39).

Furthermore, Brockriede (1970) has pointed out that people's shared sociocultural values or "norms" function as an ideology which tends to limit the attitudes of group members (p. 32). This sheds further light on students' differing abilities to write on a topic like "abortion laws." Their attitude toward these laws may be too strongly influenced by their religious, cultural, or ethnic group's norms and/or by the moral or social codes of ethics which they attribute to the perceived readers of their essays.

What kinds of personal experiences are common to most college freshmen and are not extremely biased by cultural, religious, or ethnic influences? One possible shared personal experience which might best lend itself to being shaped into the topic for a test of minimal writing competencies for CUNY freshmen is the area of education (schools, teachers, courses, skills, attendance, costs, programs, requirements, extracurricular activities, teaching machines, students' rights, etc.). Since all CUNY freshmen have gone to school for a number of years, asking them questions about this shared experience may make the test more consistent from form to form and from year to year.

Furthermore, based on experience in teaching composition courses, many teachers believe that student writing is most

fresh and lively when students write from their personal experiences and draw on material from their own lives. And, while much personal experience writing may contain narrative elements, it may also be analytic, persuasive, and explanatory. In addition, there may be an important pedagogical reason for encouraging students to use the first person pronoun in their writing: when a student uses "I," he or she is expressing his or her beliefs rather than echoing someone else's opinions and attributing them to "authorities" or to "they" or to "one."

Therefore, in this study, the independent variable called "experiential demands of the question" was operationally divided into two categories. The category called "High Experiential Demands" consisted of questions which asked students to respond with personal experiences and which used the second person pronoun, "you" in order to do this. The category which was called "Low Experiential Demands" consisted of questions which encouraged students to respond in a less personal way, with factual data and abstract generalizations.

It was hypothesized that the questions which deliberately asked for students' personal experiences would activate and sustain their writing "competence" to a much greater degree than would the questions which asked for abstractions. Moreover, it was believed that an interpretive question asking for personal experiences (i.e., a question with low cognitive

demands and high experiential demands) would be more effective in eliciting writing competence because this type of question taps their existing patterns of cognitive and affective experiences. Or put more simply, students might write better when they have something to say.

Both of the independent variables for this study were created by the investigator, and evidence that they could be validly and reliably used in an experimental study had to be provided. In order to do this, a pilot study was conducted to demonstrate that students and teachers would perceive the dimensions that were created for this study.

Sixteen questions, four replications for each of the four "treatment combination types" were submitted to four instructors and 36 students in freshman composition courses (remedial and non-remedial, native speaker and bilingual). Each question was typed on a separate page, and at the bottom of each page, respondents were asked to write answers to the following two questions: "Do you think that this question would be easy or difficult to answer in a fifty-minute essay test? Why?" ("cognitive demands"), and "What kinds of details and examples would you have to provide in order to answer this question in a fifty-minute essay test?" ("experiential demands"). There were four different orders of presentation of the 16 questions; each instructor and each group of nine students received a different order.

The results of the pilot study revealed that eight of the 16 questions were considered "difficult" by most of the students and were discarded. The other eight questions (actually four variations of two questions--one on required courses and one on the pass/fail grading system) elicited comments which corresponded to the investigator's dimensions of the two independent variables, so they were used in the study.

All four instructors and 29 students wrote that four of the eight questions discussed above were "difficult" to answer and that the other four were "easier" to answer. The four "difficult" questions corresponded to those which the investigator considered as having "high cognitive demands" and the four "easier" questions corresponded to those which the investigator considered as having "low cognitive demands."

All four instructors and 31 students agreed that four of these eight questions would require "life experiences," "personal experiences," or "things that I know about." These four questions corresponded to those which the investigator considered as having "high experiential demands." And three instructors and 30 students agreed that the other four questions corresponded to those which the investigator considered as having "low experiential demands." The eight questions which were used in this study can be found in Appendix B.

Overall Writing Quality

The main focus of the study was the effects of variation in the cognitive and experiential demands of essay questions on the overall writing quality of essays written in response to these questions. The dependent measure used to assess overall writing quality was the holistically-assigned score on the CUNY Evaluation Scale (a one-to-six scale in which six is the highest rating). Since each essay received a score from two independent readers, the dependent variable took a value from 2 (lowest) to 12 (highest).

The CUNY Evaluation Scale describes the features of writing which characterize students' essays at the 1, the 2-3, the 4-5, and the 6 levels of quality (CUNY Task Force on Writing, 1978). The scale is used to train readers in the holistic essay grading technique and to standardize their judgments of essay quality.

Basically, the holistic essay grading technique requires a subjective evaluation of an essay based on an integrated impression of its intrinsic writing characteristics. Readers are instructed not to make analytic judgments about different characteristics of an essay, but to read it quickly for an overall impression of the skill displayed by the writer (Cooper, 1977). This technique has been used for the measurement of students' writing ability and for predictions of students' success by national testing agencies such as

the College Entrance Examination Board and the Educational Testing Service, as well as by many universities and colleges (Cooper, 1971; Diederich, 1974).

The face validity of the holistic CUNY Evaluation Scale resides in the expertise of the creators of the scale. The scale was created by a committee of ten CUNY professors of English who had many years of experience in the teaching and judging of the writing of college students and who were aided in their decisions by a consultant from Educational Testing Service.

The test which they designed includes the CUNY Evaluation Scale and directions for training readers in the use of this scale. Training takes place in a three-step sequence. First, all readers must attend an intensive training session a few weeks before the actual reading. They are trained and "standardized" a second time before the first readings of actual essays and then a third time before the second readings of the essays.

The measure has high content validity because it reflects the domain of content that is considered "writing" by college English teachers. This is particularly important in light of the growing body of research which suggests that standardized, objective, machine-scorable tests of writing abilities have very low construct and content validities (Braddock, 1976; Brown, 1978; Cooper, 1978; Diederich, 1974; Hartnett, 1978;

Hooks, 1976; Odell, 1979a; Perron, 1977; White, 1978). Brown (1978) has pointed out that these tests

> leave us again with no absolute knowledge about writing ability and a slight sense of embarrassment when we tell people we'll test their writing ability by not requiring them to write a single word. Of course these tests correlate with writing ability and predict academic success; but the number of cars or television sets or bathrooms in one's family also correlate with his writing ability, and parental education is one of the best predictors there is. All existing objective tests of "writing" are very similar to I.Q. tests; even the very best of them test only reading, proofreading, editing, logic, and guessing skills. (p. 3)

The measurement of writing ability or competence can only be validly achieved by obtaining a writing sample which will be judged by human readers.

The key to the reliability of this holistic rating measure lies in the training of the scorers, described earlier. There is a body of research indicating that with carefully trained readers, interrater reliabilities range between .73 and .98 (Cooper, 1977; Diederich, 1974; Follman & Anderson, 1967; Odell, 1976; Smith, 1966). Furthermore, a holistic test of writing competence will show evidence of satisfactory validity and reliability if all passing essays meet at least a minimum standard of acceptable writing and all failing essays do not. According to a recent study of the CUNY Writing Assessment Test (Zibrin, 1979), the audit readers agreed with 84.9% of the pass/fail decisions of the original readers of 1,526 randomly sampled essays from the various CUNY colleges (p. 2). In addition, this report showed that almost 50% of the two original scorers were in exact agreement, and that

of the remaining 50%, almost 42% differed by only one point (p. 7).

Syntactic Complexity

A piece of written communication is an organic, holistic, integrated phenomenon. However, the features of written essays which may be the basic ingredients of writing fluency and may be amenable to instruction can be examined. The dependent variables in this study were selected on the basis of the relationships between these features and "overall writing quality" judgments as revealed by prior research. The feature of writing quality most frequently studied in the literature on writing performance is "syntactic complexity." As described in Hunt's (1965) pioneer study of the development of "maturity" in written composition, syntactic fluency focuses on the ways in which sentences become increasingly more elaborate through transformational processes which conjoin and embed full and reduced clauses and non-clause elements.

Hunt (1965) analyzed differences in syntactic complexity in terms of a feature he called the "minimal terminable unit" or the "T-unit." A T-unit is "a group of words which constitute a main clause in addition to all the subordinate clauses attached to it, regardless of the way the words are punctuated by the writer" (p. 19). Hunt's studies (1965, 1970, 1977) showed that mean T-unit length increases with age, and a number of other investigators have replicated his findings with students of varying ages, abilities, and sociocultural background (Blount, Johnson & Frick,

1968; Braun & Klassen 1973; Loban, 1976; Miller & Ney, 1968; Perron, 1976b; and Pope, 1974). Hunt's results have been regarded as "norms of syntactic development" for kindergarten through graduate school (Mellon, 1969; O'Hare, 1973; Stewart, 1978).

Furthermore, many researchers have found significant relationships between overall writing quality and syntactic complexity based on T-unit analyses such as mean number of T-units, mean number of words per T-unit, and percentage of multi-clause T-units (Ashida, 1977; Bateman & Zidonis, 1964; Combs, 1975; Garcia, 1977; Gebhard, 1978; Hartnett, 1978; Mellon, 1969; Najjar, 1978; O'Hare, 1973; Stewart, 1978).

One further aspect of syntactic complexity, not based on T-unit, has also been shown to be significantly related to overall writing quality: mean clause length. In her comprehensive study of writing quality and syntax, Gebhard (1978) states that "mean clause length is probably the most easily measurable indicator of writing quality available to purely quantitative assessment" (p. 217).

It is possible to measure other factors of syntactic complexity such as depth of embedding, number of embedded kernals, types of clauses, and positions of clauses. However, in his review of the different types of syntactic complexity indices, O'Donnell (1976) concluded that "T-unit length is still the most useful and useable index of syntactic development over a wide

age-range" and "clause length is the best single measure of syntactic complexity at the high school level and beyond" (p. 38).

There is one other index of syntactic complexity that has also been shown to be an important characteristic of stylistic maturity and to be significantly related to overall writing quality: the use and positioning of "free" modifiers (modifying words, phrases, and clauses set off from the rest of the sentence by punctuation). Several investigators have found significant differences in the differential use of final free modifiers (free modifiers which come at the end of main clauses) by writers of differing abilities (Dixon, 1972; Gebhard, 1978; Wolk, 1970). Furthermore, in a study of the essay characteristics which could predict readers' quality ratings, Nold and Freedman (1977) found that the number of words in free final modifiers was the second best predictor out of seventeen predictors of essay quality. Therefore, the five indices which will be used as the dependent measures of syntactic complexity are: (1) number of T-units, (2) number of words per T-unit, (3) number of clauses per T-unit, (4) number of words per clause, and (5) number of words in free final modifiers.

Sentence Control Errors

There is a small, but growing body of literature which indicates that teachers do not give a great deal of weight to syntactic fluency when they grade the writing of their students. Instead, it seems that many teachers may be more influenced by a student's ability to control sentence structure errors than by his ability to manipulate the syntactic resources of his language.

Therefore, in an attempt to provide a kind of rough balance between the assessment of syntactic complexity and the assessment of syntactic control or "correctness," this study also examined the effects of the independent variables on the sentence control errors in the essays. These errors were assessed by the following indices: (1) number of fragments per T-unit, (2) number of run-ons/comma splices per T-unit, (3) number of misplaced/dangling modifiers per T-unit, and (4) number of subject-verb agreement errors per T-unit. These four errors, as discussed and defined by Watkins, Dillingham, and Martin (1974) are shown in Table 1.

The validity of these four dependent measures comes from their basis in research which has found these indices to be significantly correlated (negatively) with writing quality (Dilworth, Jr., Reising, & Wolfe, 1978; Marzano & DiStephano, 1978; Slotnick & Rogers, 1973).

Several investigators have shown that students' ability to avoid errors in standard written English was a better predictor of teachers' quality ratings than the ability to manipulate syntactic structures (Harris, 1977; Jurgens & Griffin, 1970; Simpson, 1973; Stewart & Grobe, 1979).

In her study of high school English teachers' responses to student writing, Harris (1977) found that 66% of teachers' comments and annotations were devoted to mechanics and usage while only 12% were devoted to sentence structure (p. 179).

Table 1

List and Examples of Sentence Control Errors*

Fragment	1. A woman leaving her children at home while she works. 2. Children who have to be in daycare centers.
Run-on/ Comma Splice	1. The mother comes home from work she finds her children in trouble. 2. Women should stay home, then their entire family would be happier.
Misplaced/ Dangling Modifier	1. She went back to work after her children grew up again. 2. By going out to work, the housework never gets done.
Faulty Subject-Verb Agreement	1. Both the woman and her husband suffers. 2. My friend as well as her kids have a problem because of it.

*All examples are from actual student compositions.

And Harris concluded that while teachers state that sentence structure is an important determinant of writing quality, in actual practice, they interpret it in terms of control over errors rather than in terms of variety and appropriateness of syntactic patterns.

However, there may be a relationship between syntactic complexity and sentence control errors in that students who attempt greater sophistication in sentence structure may increase the probability of their making more errors in their sentences. Indeed, a study of the syntactic growth of college freshman writers (Maimon & Nodine, 1978) found that sentence control errors may increase as a function of developing syntactic skills.

Shaughnessy (1977) has commented on this need to examine both the syntactic style and the syntactic control of students' writing, particularly in the analysis of the writing of "basic writers" who are not at ease with formal written English. In her chapter on the syntax of 4,000 essays by college freshman basic writing students, she noted that many students "exceed the limits of their ability as writers in the effort to articulate a mature or subtle thought" (p. 46).

Vocabulary

Syntax and vocabulary are obviously interrelated: knowing a word means understanding its denotations and connotations and understanding which of its grammatical forms

is appropriate for the relationships implied by the other words in the sentence. Syntactic complexity and sentence control must be examined in light of vocabulary because, as Moffett (1968) has shown, sometimes a single well-chosen word can replace an entire clause, producing a simpler and more effective sentence (p. 174). And, as Shaughnessy (1977) has pointed out, a writer's syntactic options are often determined by the extent of his or her vocabulary (p. 73). Furthermore, there is a small body of research showing that vocabulary errors are negatively correlated with writing quality (Harris, 1977; Marzano & DiStephano, 1978; Nold & Freedman, 1977; Slotnick & Rogers, 1973).

Thus, in an attempt to extend the study of the relationships between essay questions, writing quality, and vocabulary problems, this study also examined the vocabulary of the experimental essays. This dependent measure was operationalized as one index: number of word form errors per T-unit. Examples of the various types of word form errors are listed in Table 2 (but all four types shown in this table will be added together to form one index).

Essay Length

This study also examined one further dependent variable which prior studies have shown to be related both to essay questions and to judgments of overall writing quality: essay length (the number of words per essay). Several studies have shown that essay length is related to the mode of the essay topic (Nietzke, 1972; Page, 1966; Perron, 1976a, 1976b; Rosen, 1969; Stallard, 1974).

Table 2

Types and Examples of Word Form Errors

Non-Existent Form*	She wanted to be securitated.
Form With an Incorrect Affix	She could not work without supervise.
Phonetically Similar Form Which is Semantically Incorrect	He wanted her to work in that pacific area.
Form Which Violates Semantic Constraints of The Sentences	By working, a woman's financial status is enlightened.
Form Which is Idiomatically Incorrect	They are working regardless to their husbands' decisions.

*When a word is misspelled by more than two letters, it will be considered a non-existent form (Shaughnessy, 1977, p. 190).

METHODOLOGY

Research Design

This research was a quasi-experimental study of the effects of two essay question variables on students' writing performance. It was "quasi" because full experimental control could not be achieved (Campbell & Stanley, 1966). The most appropriate research design for investigating and for statistically testing the interactive effects of two independent variables on a dependent variable is a factorial analysis of variance (ANOVA) approach (Kerlinger, 1973); Lindquist, 1953; Winer, 1962). ANOVA techniques enable a researcher to manipulate and control both independent variables simultaneously and to hypothesize about and test for the interaction of those variables.

In order to separate out the effects of each independent variable, this approach requires that the independent variables be unrelated. In order to ensure this, subjects were randomly assigned to the different treatment combination cells of the independent variables, and there was the same number of observations in each cell (Winer, 1962).

For this study, each independent variable was partitioned into two categories: "high" and "low." This created four treatment combinations which were examined in a "Two-by-Two Factorial Experiment" (Winer, 1962). This design eliminated

sequence effects because each subject wrote only on one treatment combination (one type of question). While this design could not control the effects of individual differences, it did randomize these effects insofar as these were equalized by the random assignment of subjects to treatments and by the relatively large sample size. (This design is illustrated in Figure 1).

Sample

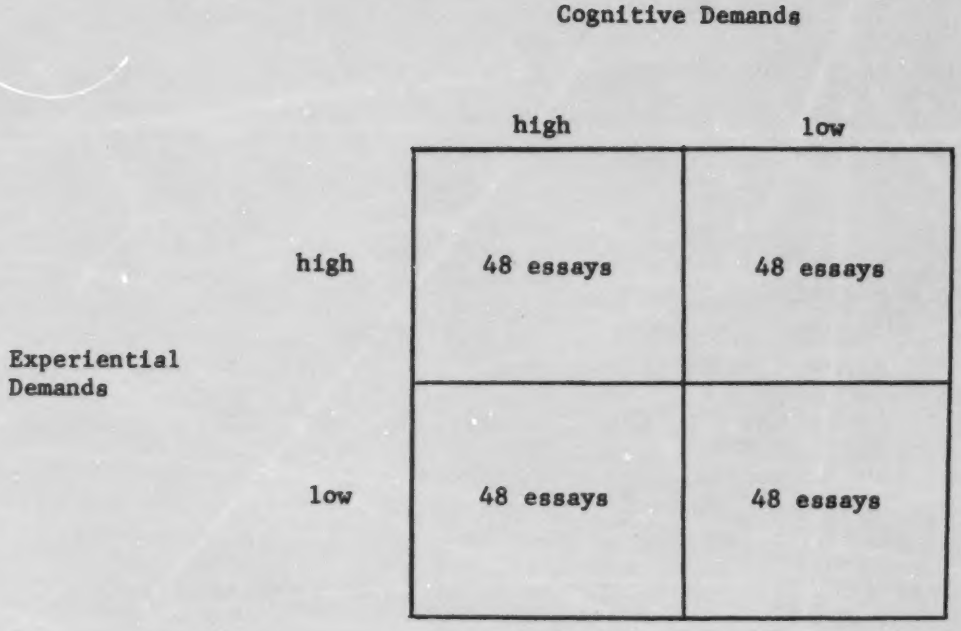
The study design described above required 192 essays: 48 subjects writing essays on each of the four types of questions. All 192 students were selected from 12 classes which participated in the study: two lower-remedial native speaker classes, two lower-remedial English as a Second Language (E.S.L.) classes, two upper-remedial native speaker classes, two upper-remedial E.S.L. classes, and four non-remedial (college-level freshman composition) classes. Thus, the sample was representative of the population. Within each of the 12 classes, 16 essays were randomly assigned for experimental analysis (four essays on each of the four types of questions). With 192 subjects (48 per cell), the "power" of a significant "medium" interaction effect is .74 (Cohen, 1969).

Experimental Treatment

The four combinations of the two independent variables which were experimentally manipulated in this study were as follows:

Figure 1

Research Design*



*Each cell contains 48 students.

8 low-remedial, native English speakers

8 upper-remedial, native English speakers

8 lower-remedial, E.S.L. speakers

8 upper-remedial, E.S.L. speakers

16 non-remedial students

(A) Cognitive demands of an essay test question:

(a_1) High: students were asked to evaluate an issue by agreeing or disagreeing with a statement about the issue.

(a_2) Low: students were asked to interpret an issue by discussing a statement about the issue.

(B) Experiential demands of an essay test question:

(b_1) High: students were asked to respond with personal experiences.

(b_2) Low: students were asked to respond with generalizations and facts.

Dependent Variables

Overall Writing Quality

Overall writing quality was assessed by the holistically-assigned score on the CUNY Evaluation Scale (described in the section on Measurement and Operational Definitions). The scale is a six-point scale (with 6 being the highest score). Because each essay is scored twice with this scale (by two different readers), an essay's final score can range from 2 through 12.

The holistic scoring of all 192 essays was done in a single session at the yearly audit of the WSAT conducted at the Central Office of the University. Twenty-two experienced CUNY Writing Assessment Test readers attended a training session in holistic scoring. After training, but prior to

the scoring of experimental essays, inter-rater reliability coefficients were calculated (using sample data) by means of Ebel's (1951) formula for determining the reliability of a mean rating from a group of raters (p. 423). After satisfactory reliability (above .87) was achieved, the readers independently read and scored the experimental essays which were randomly mixed in with the 1700 audit essays. After each essay was scored once, a tape was placed over the score so that it would not influence the second reader.

The training and first reading/scoring of essays took place from 9 A.M. to 1 P.M. Then there was a one hour long lunch break which also served as an hour long rest period to avoid the erratic evaluations which result from reader fatigue (Braddock, Lloyd-Jones, & Schoer, 1963, Godshalk, Swineford, & Coffman, 1963). The second reading/scoring of the essays took place from 2 P.M. to 5 P.M. of the same day.

Syntactic Complexity, Sentence Control Errors, Word Form Errors, and Essay Length

One hundred fifty-two essays were randomly selected and were coded for the other dependent variables (syntactic complexity, sentence control errors, word form errors, and essay length), with 152 subjects (38 per cell), the "power" of a significant "medium" interaction effect is .68 (Cohen, 1969).

Six graduate students were hired and trained to code these dependent variables, and they coded the essays for each dependent variable separately.

Procedures for segmenting and counting the T-units were those listed by O'Hare (1973, pp. 47-49), and procedures for counting free final modifiers were those listed by Wolk (1970, pp. 61-62). The six coders were trained according to these procedures using non-experimental essays. The inter-coder reliabilities for each dependent variable can be found in Appendix C (all but "number of subordinate clauses" were .90 or higher; "number of subordinate clauses" was .83).

The coders worked independently on xerox copies of the essays without grades, other coding, or notations of any kind on them. All of the coders worked in the same room with the investigator present to answer any questions. There were two coding sessions: from 9 A.M. to 1:30 P.M. and from 2:30 P.M. to 6:00 P.M.

Covariates

Any significant interaction effects were to be adjusted by an analysis of covariance which would control for the variance in the dependent measure which was due to individual differences in students' reading ability and native language.

These two covariates were selected as control variables because they may affect students' writing performance (R. Brown, 1978; Dilworth, Jr., Reising, & Wolfe, 1978; Nietzke, 1971; Plec, 1978; Rothkopf & Thurner, 1970).

The reading ability of all 192 students was assessed by the Descriptive Tests of Language Skills (DTLS) subtest of reading comprehension skills, Forms A and D (College Entrance Examination Board, 1970). This test measures a student's ability to recognize main ideas, understand direct statements, and draw inferences. The raw score range is from 1 through 45.

All of the students took the DTLS when they entered the college and their scores were collected from their transcripts. Thus, there is no reliability figure for this sample of students. According to ETS, however, the Kuder-Richardson-20 reliability coefficient for the DTLS reading comprehension test is .89 and the standard error is 2.1 (Educational Testing Service, 1978).

Information on the first/native language of all 192 students was also collected from the students' transcripts. This covariate was dichotomized as "English" or "other," and was "dummy coded" as 1 or 0 for statistical analysis.

Procedures

Teachers were asked to administer all of the questions to each class on the same day in a 50-minute class. The letter to the faculty is in Appendix B. All students wrote their essays in the same week.

All of the questions were typed on the same paper that is used for actual CUNY Writing Assessment Tests, and students were given genuine test booklets in which to write their essays. Teachers were instructed to tell the students to put their names on the backs of the booklets, but only their social security numbers inside. The investigator picked up, xeroxed, and returned the essays to the teachers on the same day that they were written.

Supplementary Analyses

In order to determine whether essays assigned failing holistic quality scores (2 through 7) and essays assigned passing holistic quality scores (8 through 12) differed in their mean syntactic complexity, mean number of sentence control errors, mean number of word form errors, and mean number of words, eleven separate t-tests were performed on the eleven selected essay characteristics.

In order to determine the extent of the relationship between the essays' holistic quality scores and the quantitatively measured essay charactieristics, a bivariate correlational analysis and a stepwise multiple regressional analysis were performed.

RESULTS

The mean, the standard deviation, the mode, and the median for the twelve dependent variables and the two continuous covariates in this study are presented in Table 3.

The mean holistic essay score or quality rating (QWQ) of 6.92 is very close to the arithemtic mean (7) of the full range of scores between 2 and 12. However, the most prevalent score was 6: 45 of the 192 essays were assigned a 6 by two or more of the readers. The distribution of quality scores in this sample is almost symmetric. It is slightly positively skewed because five essays received a score of 12, but no essay received a score of 2. The standard deviation is 1.91; 80% of the scores fell between a 5 and a 9.

In this distribution of quality scores, both the mean and the median are below the score of 8, which is the score mandated by the University as the minimum for exemption from remedial writing courses. Students who score between 8 and 12 are not assigned to remediation in writing at most of the unit colleges.

However, in this sample, two-thirds of the students already were in remedial writing classes, and an examination of the mean quality score for these students (6.41) and the

Table 3

Means, Standard Deviations, Modes, and Medians

for the Dependent Variables

N = 192	Mean	SD	Mode	Median
Holistic score (OWQ)	6.92	1.91	6.00	6.74

N = 152	Mean	SD	Mode	Median
Holistic score (OWQ)	6.81	1.83	6.00	6.64
No. of T-units (TUS)	18.55	5.80	16.00	18.67
No. of words per T-unit (WDSTU)	17.61	4.26	16.25	16.79
No. of clauses per T-unit (CLSTU)	1.55	0.29	1.50	1.50
No. of words per clause (WDSCL)	11.46	2.71	10.40	11.13
No. of words in free final modifiers (WDSFFM)	5.23	11.89	0.00	0.32
No. of fragments per T-unit (FRAGSTU)	0.05	0.07	0.00	0.00
No. of run-ons per T-unit (ROSTU)	0.05	0.07	0.00	0.00
No. of misplaced/dangling modifiers per T-unit (MMSTU)	0.03	0.05	0.00	0.00
No. of subject-verb agreement errors per T-unit (SVAESTU)	0.06	0.13	0.00	0.04
No. of word form errors per T-unit (WFESTU)	0.48	0.41	0.00	0.36
No. of words per essay (WDS)	313.60	88.81	260.00	312.50

Table 4

A Comparison of Mean Holistic Scores (OWQ)
for Students in Remedial Writing Courses and
for Students Not in Remedial Writing Courses

Group	N	Mean	SD	SE	t	df
Remedial	128	6.414	1.834	0.162		
					-5.60*	190
Non-Remedial	64	7.938	1.661	0.208		

*p ≤ .001

Table 5

A Comparison of Mean Holistic Scores (OWQ)
for Native English Students and
for English-as-a-Second Language Students

Group	N	Mean	SD	SE	t	df
English	128	7.500	1.861	0.165		
					6.53*	190
E.S.L.	64	5.766	1.445	0.181		

*p ≤ .001

mean quality score for the non-remedial students in this sample (7.94) revealed that the means of these two subsamples were significantly different (Table 4).

The distribution of scores for the entire sample was skewed negatively toward scores below the CUNY passing score of 8 in large part because there was a large proportion of remedial writing students in this sample.

The other factor which accounts for this result is the difference in the mean quality scores between students whose native language is English (7.50) and students who are learning English as a new language (5.77). The mean quality scores of these two subsamples also differ significantly (Table 5).

While the mean holistic quality score of the essays of the entire sample (6.92) was not superior, the mean number of words per T-unit (17.61) was surprising. Gebhard (1978) has reported a mean of 15.37 words per T-unit for "poor" writers (at the college freshman level) and a mean of 16.33 words per T-unit for "good" writers. And Hunt (1965) has reported 14.4 words per T-unit at grade 12 and 20.3 for superior adults. In this study's sample, 18% of the essays had an average number of 20 or more words per T-unit.

However, this sample did not do as well in terms of mean number of words per clause and words in free final

modifiers. Gebhard (1978) reported a mean of 10.28 words per clause for poor freshman writers and 11.60 for good writers. This sample averaged only 11.46 words per clause. However, it must be pointed out that Gebhard's and Hunt's samples did not have as many ESL students as did this sample, and this fact may account for some of the results presented above.

Another interesting and unexpected result of the descriptive analyses of the data was the remarkably low frequency of sentence control errors in this sample's essays. Not one essay had more than five fragments, and the mean number of fragments per T-unit was very low: 0.05. Fifty-four percent of the essays had no fragments at all.

In addition, no essay had more than seven run-ons, and 58% had none. No essay had more than five misplaced and/or dangling modifiers, and 66% had none. This lack of sentence control errors seems remarkable in light of the fact that two-thirds of the students required remedial help with their writing.

However, these students did have many problems with vocabulary and diction. Forty-five percent of the students wrote eight or more word form errors per essay, nine percent wrote twenty or more per essay, and the mean number of word form errors per T-unit was 0.48.

Another interesting result concerned the number of words that the students in this sample wrote. All eight experimental questions asked for "an essay of about 300 words," and the mean number of words per essay was 313.6. Fifty-four percent of the students wrote more than 300 words per essay, and sixteen percent wrote more than 400 words per essay.

Table 6 gives the simple correlations among all of the independent variables, the dependent variables, and the covariates. A minus sign means that the two variables are negatively correlated. The correlations give the first hint about the results of the experiment: holistic score (OWQ) is not significantly correlated with either of the independent variables (cognitive and experiential demand levels of the questions).

Table 6

Simple Correlations for all Variables

N = 152

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Holistic Score (OWA)																
2 Topic	-.09															
3 Exp. Demands	-.15*	.09														
4 Cog. Demands	-.09	.99**	.09													
5 DTLS	.43***	-.12*	-.05	-.12												
6 Native Lang	-.43***	.02	.00	.02	-.42***											
7 No. Words	.19**	-.18*	-.33*	-.18*	.20**	-.03										
8 No. T-units	.06	-.17*	-.18*	-.17*	.02	.03	.67*									
9 No. Words/TU	.17*	.00	-.05	-.01	.24**	-.-9	.19	-.53***								
10 No. Clause/TU	-.02	.15*	-.09	.16	.01	.13	.17*	-.28***	.59***							
11 No. Words/Clause	.19**	-.14*	.04	-.16*	.25***	-.21*	.02	-.39***	.63***	-.23**						
12 No. Words/FFM	-.15*	.14	.11	.13	-.02	-.01	.01	-.01	-.01	.00	-.01					
13 No. Frags./TU	-.24**	-.03	.07	-.01	-.17*	.13	-.24***	-.22**	.05	.07	.05	-.09				
14 No. Run-ons/TU	-.23**	-.01	.09	.01	-.16*	.28***	.02	-.01*	.06	.13	-.05	.27***	-.01			
15 No. MMs/TU	.02	-.07	.12	-.09	.15	-.14*	.07	-.17*	.38***	.09	.32***	.06	-.07	.19**		
16 No. SubVerb Errors/TU	-.30***	.07	.22**	.06	-.12	.14	-.18*	-.26***	.20**	.28***	-.04	.03	.10	.26***	.13	
17 No. Word Errors/TU	-.47***	.07	-.01	.06	-.03***	.34***	-.19*	-.22**	.10	.09	.07	.03	.17	.24***	.15	.20**

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

Table 7

A Comparison of Mean Holistic Scores (OWQ)

for the Different Topics

Group	N	Mean	SD	SE	t	df	Sig.
Grading Systems	96	6.979	2.010	0.204			
Required Courses	96	6.863	1.843	0.189	0.42	190	N.S.

Table 8

An Analysis of Variance

for Mean Holistic Scores (OWQ)

N = 192

Source	df	MS	F	Sig.
Main Effects	2			
COG	1	2.068	0.557	N.S.
EXP	1	6.968	1.877	N.S.
Interactions				
COG x EXP	1	0.528	0.142	N.S.
Residual	188	3.706		

Table 9

An Analysis of Variance

for Mean Number of Words in Free Final Modifiers

$N = 152$

Source	df	MS	F	Sig.
Main Effects				
COG	1	264.145	2.003	N.S.
EXP	1	183.300	1.390	N.S.
Interactions				
COG x EXP	1	1096.073	8.312	$p \leq .01$
Residual	148	131.873		

Table 10

An Analysis of Variance

for Mean Number of T-units

N = 152

Sources	df	MS	F	Sig.
Main Effects				
COG	1	90.250	2.771	$p \leq .05$
EXP	1	121.803	3.740	$p \leq .05$
Interactions				
COG x EXP	1	2.143	0.066	N.S.
Residual	148	32.570		

Table 11

An Analysis of Variance

for Mean Number of Clauses per T-Unit

N = 152

Source	df	MS	F	Sig.
Main Effects				
COG	1	0.400	4.908	$p \leq .05$
EXP	1	0.150	1.840	N.S.
Interactions				
COG x EXP	1	0.001	0.012	N.S.
Residual	148	0.081		

Table 12

An Analysis of Variance

for Mean Number of Words per Clause

N = 152

Source	df	MS	F	Sig.
Main Effects				
COG	1	27.734	4.178	$p \leq .05$
EXP	1	3.966	0.597	N.S.
Interactions				
COG x EXP	1	23.076	3.476	N.S.
Residual	148	7.047		

Table 13

An Analysis of Variance

For Mean Number of Words per Essay

N = 152

Source	df	MS	F	Sig.
Main Effects				
COG	1	24106.707	3.551	$p \leq .01$
EXP	1	106013.875	15.441	$p \leq .001$
Interactions				
COG x EXP	1	18235.613	2.656	N.S.
Residual	148	6865.609		

Table 14

An Analysis of Variance

for Mean Number of "I's" per T-Unit

N = 152

Source	df	MS	F	Sig.
Main Effects				
COG	1	0.435	6.337	$p \leq .05$
EXP	1	0.006	0.089	N.S.
Interactions				
COG x EXP	1	0.262	3.817	N.S.
Residual	148	0.069		

Table 15

A Comparison of Mean Scores

on Eleven Essay Characteristics

for Passing and Failing Essays

Variable	Group[1]	N	Mean	SD	SE	t	df
WDS							
	F	100	307.43	85.31	8.53	-1.21	150
	P	52	325.69	93.58	12.98		
TUS							
	F	100	18.58	5.69	0.58	-0.02	150
	P	52	18.59	6.03	0.84		
WDSTU							
	F	100	17.16	3.96	0.39	-1.70	150
	P	52	18.38	4.68	0.65		
CLSTU							
	F	100	1.57	0.30	0.30	1.16	150
	P	52	1.52	0.26	0.38		
WDSFFM							
	F	100	6.20	13.72	1.38	1.51	150
	P	52	3.15	6.51	0.90		
WDSCL							
	F	100	10.97	1.89	0.19	-2.97**	150
	P	52	12.31	3.68	0.51		
FRAGSTU							
	F	100	0.06	0.08	0.01	2.34*	150
	P	52	0.03	0.05	0.01		
ROSTU							
	F	100	0.54	0.08	0.01	1.86*	150
	P	52	0.03	0.06	0.01		
MMSTU							
	F	100	0.03	0.04	0.00	-0.28	150
	P	52	0.03	0.06	0.01		
SVAESTU							
	F	100	0.10	0.12	0.01	2.13*	150
	P	52	0.05	0.14	0.02		
WFESTU							
	F	100	0.54	0.43	0.04	5.27***	150
	P	52	0.25	0.25	0.03		

1 F denotes failing essays (OWQ = 2-7)
P denotes passing essays (OWQ = 8-12)

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

Table 16

A Summary Table of a Stepwise Multiple Regression of Mean Holistic Score (OWQ) and Possible Predictor Variables (Essay Characteristics)[1]

Step Number	Variable	Multiple R^2	Multiple R^2	Multiple R^2 Change	B	Standard Error
1	Word form errors	0.455	0.207	0.207	-0.113	0.019
2	Words	0.508	0.258	0.051	0.841	0.002
3	Fragments	0.549	0.302	0.044	-0.317	0.111
4	Subject-verb agreement errors	0.579	0.336	0.034	-0.188	0.076
5	Words in free final modifiers	0.598	0.358	0.022	-0.214	0.010
6	Run-ons	0.613	0.375	0.017	-0.356	0.157
7	Subordinate Clauses	0.620	0.385	0.010	-0.567	0.033
8	T-units	0.624	0.389	0.004	-0.303	0.030

1 df = 150
2 F value to enter a variable into the equation = p ≤.01

The correlations also show that the essay characteristics most significantly correlated with holistic score (OWQ) are subject-verb agreement errors per T-unit (r = -.30, $p \leq .001$) and word form errors per T-unit (r = -.47, $p \leq .001$) both of which are negatively correlated with holistic scores.

Furthermore, the correlations reveal that the question topic, ("the pass/fail grading system" or "required courses"), was not significantly correlated with holistic score. A T-test on the differences between the two topics indicated that they were not significantly different (Table 7). This finding lends support to the assumption that the two topics of the different experimental questions were similar.

In order to determine the interactive effects of the two independent variables (cognitive demands and experiential demands of the question) on each of the dependent variables, a separate two-way factorial analysis of variance was conducted for each of the dependent variables.

Hypothesis One

There will be a significant interaction between the cognitive and the experiential demand levels of the essay questions with Low Cognitive/High Experiential questions eliciting higher overall writing quality.

Hypothesis One was not supported. A .05 level test of the significance of the interactive effects of cognitive demands and experiential demands failed to demonstrate a statistically significant relationship between the interaction of these factors and holistic essay grades (Table 8). Thus, no analyses of covariance were done.

Hypothesis Two

There will be a significant interaction between the cognitive and the experiential demand levels of the essay questions with Low Cognitive/High Experiential questions eliciting greater syntactic complexity, fewer sentence control errors, fewer word form errors, and longer essays.

Hypothesis Two was not supported either. Eleven .05 level tests of significance of the interactive effects of cognitive demands and experiential demands failed to demonstrate a statistically significant relationship between the interaction of these factors and ten of the eleven dependent variables. The one dependent variable which was significantly related to the interaction of the independent variables was the mean number of words in free final modifiers (Table 9).

Although Hypothesis Two was rejected because of the nonsignificant interaction effects of the independent variables on the dependent variables, the ANOVA's revealed significant relationships between the independent variables and four of the dependent variables.

The mean number of T-units per essay (18.55) was significantly affected by the cognitive dimensions of the questions and by the experiential dimensions of the question but not by the interaction of these dimensions (Table 10). Students who wrote about questions with high cognitive demands wrote an average of 1.86 more T-units per essay than those who wrote about questions with low cognitive demands. And students who wrote about questions with high experiential demands wrote an average of 2.12 more T-units per essay than students who wrote about questions with low experiential demands.

In addition, the main effect of the independent variable "cognitive demand level" had a significant effect on the mean number of clauses per T-unit and words per clause (Table 11 and 12). Students who wrote about questions with low cognitive demands wrote an average of 0.10 more clauses per T-unit and an average of 0.86 more words per clause than students who wrote about questions with high cognitive demands.

The mean number of words per essay was significantly affected by both the cognitive and the experiential demand-levels of the questions (Table 13). Questions with high cognitive demands elicited an average of 32.24 more words per essay than did questions with low cognitive demands. And questions with high experiential demands elicited an average of 58 more words per essay than did questions with low experiential demands.

Another surprising result pertains to a dependent variable which was not included in the original set of hypotheses: the number of "I's" (first person singular pronouns) per T-unit. The frequency of this variable was also counted by the coders (with an inter-coder reliability of .95) because of an interest in determining whether questions with high experiential demands (which deliberately used the second person pronoun, "you," in order to ask for personal experiences) would elicit more "I's" in response to them. The mean number of "I's" per T-unit was 0.20, and an ANOVA showed that the experiential demand level of the questions had no significant effect on the number of "I's" per T-unit. Students who answered questions which specifically asked them to write about themselves wrote an average of only 0.02 more "I's" per T-unit than did students who answered the other types of questions (Table 14).

This may explain another surprising result: the number of "I's" per T-unit that the students wrote were not significantly related to the essays assigned holistic score.

Hypothesis Three

There will be a significant difference in the syntactic complexity, the number of sentence control errors, the number of word form errors, and the number of words between essays assigned failing holistic quality scores (2 through 7) and essays assigned passing holistic quality scores (8 through 12).

Hypothesis Three was supported in part. Five t-tests of the significance in the differences of the means of the passing and the failing essays on the selected essay characteristics revealed that there were no significant differences between passing and failing essays in their syntactic complexity (mean number of T-units, words per T-unit, clauses per T-unit, and words in free final modifiers) and in their length (mean number of words) (Table 15).

However, four t-tests of the significance in the differences of the means of the passing and failing essays showed that there were significant differences in the number of sentence control errors (mean number of fragments per T-units, run-ons per T-unit, and subject-verb agreement errors per T-unit) and in the mean number of word form errors (Table 15).

Hypothesis Four

The essays assigned holistic quality scores will be significantly related to the quantitatively measured characteristics (of syntactic complexity, number of sentence control errors, number of word form errors, and number of words).

Hypothesis Four was supported in part. A bivariate correlational analysis showed that the essays' holistic scores were

significantly related to the number of fragments per T-unit, the number of run-ons per T-unit, the number of subject-verb agreement errors per T-unit, and the number of word form errors per T-unit (Table 6).

In order to determine the extent to which the essays' holistically-assigned quality scores could be predicted by all of the quantitatively measured essay characteristics, a stepwise multiple regressional analysis was performed. This analysis showed that, in combination, eight essay characteristics accounted for 38.93% of the variance in the essays' holistically-assigned quality scores (Table 16). Of this 38.93% almost 21% (20.7%) of the variance in holistic scores was accounted for by the number of word form errors per T-unit in the essays.

DISCUSSION AND CONCLUSIONS

The central experiment in this study hypothesized that interpretive essay questions which offered students a variety of options for structuring answers (i.e., had "low cognitive demands") and which asked students for their personal experiences (i.e., had "high experiential demands") would improve the quality of students' written responses to these questions.

The results of this study did not confirm this hypothesis. Students' overall writing quality, as measured by assigned holistic essay score, was not significantly affected by the type of essay question to which they responded. The findings of this study showed that none of the four types of questions elicited a substantially superior measure of rhetorical, syntactic, or lexical performance.

The results did disclose the following statistically minor effects of the question dimensions on the selected essay characteristics. "High cognitive" questions elicited more words and T-units per essay, but "low cognitive" questions elicited more clauses per T-unit, words per clause, and words in free final modifiers. And "high cognitive" questions elicited more words and T-units per essay and more words in free final modifiers. The interaction of the cognitive and experiential dimensions significantly affected only the number of words in free final modifiers: "low cognitive/high experiential" questions elicited more words in free final modifiers than did any other type of question. However, the interaction of the cognitive and experiential demands did not significantly affect the quality or fluency of these students' essays.

It was surprising to find that changes in the cognitive and experiential demands of essay questions produced no significant

measurable changes in students' writing performance. Equally surprising was the finding that the "high experiential" questions which used the second person pronoun, "you," to ask students to discuss their personal experiences not only did not elicit better writing, but did not even elicit a higher frequency of first person "I's" in response.

Perhaps students are resistant to using the first person pronoun in their essays because they or their instructors believe that it is inappropriate for "formal" essays. Composition textbooks and handbooks warn students to "use only accurate and verified data," "use reliable authorities," and "use specific evidence" (Watkins, Dillingham, & Martin, 1974, pp. 185-187). Students may have come to accept these commands and do not consider themselves to be "reliable authorities" with opinions worth defending and with experiences worth writing about. It may be safer for them to write about "one" or "students," particularly if they are echoing someone else's ideas or are writing to please their perceived readers.

However, just because the students in this sample did not respond to the "you" in these test questions, it does not necessarily follow that they would not respond in other writing tasks and contexts. Indeed, personal writing is an effective teaching tool for students in "basic" and "remedial" writing courses because it gives them topics on which they can elaborate, and it allows them to work out patterns of development which will effectively communicate their experiences. Many classroom teachers believe that personal writing enlivens all types of essays--narrative, analysis, and argument.

In addition to students' resistance to incorporating personal experience in these experimental essays, there may be other reasons why the change in the wording and in the demands of the questions had no significant effects on students' writing performance. It is quite possible that the anxiety which accompanies a timed, in-class essay test is so great for so many students that they cannot fully comprehend the directions of the writing task and do not take advantage of the information contained in or the strategies offered by the questions.

In fact, many of the students in this sample were "inexperienced" writers who were in or who had been in remedial writing courses. They may still be too insecure and anxious about writing to attend to variations in question wording which would affect more sophisticated, skillful student-writers.

In addition to the unexpected lack of significant effects of variations in this study's questions on students' writing performance, the study produced other surprising findings. One particularly unexpected result concerned the types of syntactic and semantic cues within essays which are related to readers' responses to and judgments of the writing.

More extensive differences might have been expected to be found in this sample than were actually evidenced because these differences were discovered in so many other studies (Christensen, 1963; Garcia, 1977; Gebhard, 1978; Hartnett, 1978; Hunt, 1965, 1970, 1977; Nold & Freedman, 1977; O'Donnell, 1976; O'Hare, 1973; Stewart, 1978; and Wolk, 1970). However, the samples in these studies did not contain as many ESL students as did the sample in the current study.

The passing and failing essays in this study did not differ significantly on four major indices of syntactic complexity: number of T-units per essay, words per T-unit, clauses per T-unit, and number of words in free final modifiers. However, the essay features which did significantly distinguish the passing from the failing essays were the types and frequencies of errors: number of fragments, run-ons, subject-verb agreement errors, and word form errors.

The above findings are interesting in light of the fact that, as a sample, these students' essays showed remarkably low frequencies of errors. More than half of the essays had no fragments, run-ons, or misplaced modifiers. In addition, almost half of these students did not write any subject-verb agreement errors and only one quarter of them wrote more than one. This low frequency of subject-verb agreement errors is especially surprising considering that one third of this sample is learning English as a new language.

Thus, most of the students in this sample did not write many sentence structure errors or subject-verb agreement errors, yet the ones which they did write seemed to have had a disproportionately large influence on the readers' perceptions of the quality of the essays. Assigned holistic essay score (OWQ) was negatively correlated with number of fragments per T-unit ($\underline{r} = -.24$, $\underline{p} \leq .01$), run-ons per t-unit ($\underline{r} = -.30$, $\underline{p} \leq .001$).

The above finding confirms the findings of other researchers who have shown that students' ability to avoid errors in standard written English is a better predictor of teachers' quality ratings than the ability to manipulate complex syntactic structures (Harris, 1977; Jurgens & Griffin, 1970; Simpson, 1973; Stewart & Grobe, 1979).

The major problems exhibited in these students' essays were word form errors. Almost half the sample wrote eight or more word form errors per essay. All of the word form errors in this study were coded as one variable (categories of word form errors were not identified or coded), but an examination of the 1,145 errors indicated that there were several different types of word form errors.

There were many idiomatically incorrect forms, particularly incorrect preposition substitutions. There were also many words with incorrect affixes: "It would not be fair for a person with a high average to be grading with a pass/fail system." For these forms, there is no way of knowing whether the student switched suffixes carelessly or is still unsure about the use of different suffixes.

Another memorable type of word form error was the one in which the student produced a form which sounded familiar, but which did not exist: "Firstable, the grading system is fair." And, finally, there were many seemingly careless errors and spelling errors: "I have almost fishised my freshman year at College."

Shaughnessy (1977) has said that all of these word form errors are "advanced" errors because they are signs of students' attempts to acquire formal academic vocabulary, whatever the cost in error. Given the fact that these students wrote relatively long essays and T-units, the high frequency of these word form errors may indeed be an indication that many of these students are trying to practice a new, more formal style and are stumbling over some of their "academese."

Unfortunately for these students, however, their imprecise or incorrect word forms did affect teachers' evaluations of their writing. The number of word form errors was the only essay characteristic to be strongly correlated with assigned holistic score and the correlation was negative ($\underline{r} = -.47$, $\underline{p} \leq .001$). Furthermore, the number of word form errors per T-unit in an essay was the best predictor of the essay's assigned quality score, accounting for 20.7% of the variance in assigned holistic scores.

This finding suggests that the readers of this type of test of minimal writing competencies are quite sensitive to lexical dissonance. Of course, this is to be expected because lexical errors quite frequently produce syntactic dissonance and blurred meaning.

In the final analysis, it seems that the results of this study demonstrate the complex, integrated nature of all of the language processes. While the writing process may be influenced by changes in the format, content, wording, and context of the writing assignment, students' fluency in and control of the written dialect seem to be more dependent on the level of development of their oral and reading abilities. Students who are more comfortable with and fluent in the academic written dialect generally produce better essays regardless of the writing task, topic, or question.

The findings of this study showed that changes in the cognitive and experiential demands of essay test questions did not affect the test's potential to elicit students' minimal writing competencies. Offering students a variety of options for answering the questions had no effect of the quality of students' answers, nor did asking students to relate the questions to their own personal

experiences. These results, then provide further empirical construct validation for the CUNY Writing Skills Assissment Test (WSAT).

In summary, this study experimentally manipulated two aspects of the WSAT questions (the "cognitive demands" and the "experiential demands") in an attempt to examine whether these variations would increase the test's potential to elicit students' minimal writing competencies. The results showed that neither of the changes had any significant effects on students' writing performance. The implication of these results is that the questions on the WSAT are effective stimuli for eliciting minimal writing competencies.

In his evaluation of the WSAT, Odell (1979a) stressed that

> designing and testing new writing assignments, and analyzing sample essays so as to refine, where necessary, both writing assignments and the criteria by which student essays are scored . . . are essential to maintaining the validity of the CUNY Writing Skills Assessment Test. (pp. 7-8)

The results of this study provide a step in this direction by validating the format of the WSAT's writing assignments.

Obviously, there is much more research that needs to be conducted on the WSAT. Studies similar to this one should be carried out to test the effects of variations in other characteristics of the WSAT questions: number of questions per test, number of words per questions, levels of semantic difficulty, levels of syntactic difficulty, levels of rhetorical difficulty, etc. Also, this study should be replicated with different types of populations at other branches of CUNY.

PEDAGOGICAL IMPLICATIONS

Writing is a dynamic complex of skills, processes, behaviors, and abilities that develops with time, maturity, practice, and teaching.

Unfortunately, no one has identified writing tasks or topics which consistently stimulate or elicit superior writing from students. Thus, teachers must continue to find or to formulate topics, tasks, and questions which motivate their students and facilitate the students' writing improvement.

While the findings of the experiment in this study showed that changes in essay test questions did not affect students' writing performance, these findings do not imply that teachers should not devote a great deal of time and ingenuity to the creation of interesting, challenging writing tasks in a variety of modes on a variety of topics for their students. The only consistent characteristic of the passing essays was fluency in written language, and this characteristic cannot be mastered by practicing any one particular mode, format, or topic. Probably the best way to help students become more fluent is to have them experiment with a wide variety of modes, formats, and topics and to help them to vary their rhetorical and syntactic strategies and their vocabulary to fit different purposes and readers.

It is important to remember that the findings of this study concern the structuring of writing tasks for tests of minimal writing compentencies, not for structuring in-class or out-of-class writing assignments. It may be that the experimental variables in this study--the levels of cognitive and experiential demands--do influence students' written rhetorical, syntactic, and lexical

performance, but these effects may not show up in a "test" situation. Students may need more time and/or more opportunities for discussion in order to effectively respond to the different strategies offered in the questions.

There is a need for descriptive studies of the teaching and the learning processes which occur in writing courses. These studies should examine the effects of different types of writing tasks, assignments, questions, and the feedback which students receive from teachers on their writing to discover how these are related to the improvement of student writing. Teachers and testmakers must continue to search for stimuli which promote cognitive and linguistic growth and development.

Finally, if we are going to effect significant and lasting changes in students' writing competence, we must look more carefully at our students' perceptions of and cognitive/affective responses to our writing assignments.

REFERENCES

Ashida, M.E. Form, syntax, and statistics: A quantitative approach to written composition. Unpublished doctoral dissertation, The University of Nebraska, 1967.

Bateman, D.R., & Zidonis, F.J. The effects of a study of transformational grammar on the writing of ninth and tenth graders. Champaign, Illinois: National Council of Teachers of English, 1974.

Belanger, J.F. Calculating the syntactic density score: A mathematical problem. Research in the Teaching of English, 1978, 12, 149-154.

Bloom, B.S. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: Longmans, Green, 1956.

Blount, N.S., Johnson, S.L. & Frick, W.C. A comparison of the writing of eighth- and twelfth-grade students. Technical Report No. 78. Madison, Wisconsin: University of Wisconsin Research and Development Center for Cognitive Learning, 1968. (ERIC Document Reproduction Service No. ED 035 652).

Braddock, R. Evaluation of writing tests. In A.H. Grommon, R. Braddock, J.N. Hook, W.A. Jenkins, W. Loban, & A.C. Purves (Eds.), Review of selected published tests in English. Urbana, Illinois: National Council of Teachers of English, 1976.

Braddock, R., Lloyd-Jones, R. & Schoer, L. Research in written composition. Champaign, Illinois: National Council of Teachers of English, 1963.

Braun, C., & Klassen, B. A transformational analysis of written structures of children representing varying ethno-linguistic communities. Research in the Teaching of English, 1973, 7, 312-323.

Britton, J. Language and learning. Baltimore, Maryland: Penguin Books, 1970.

Britton, J., Martin, N., & Rosen, H. Multiple marking of compositions. London: Her Majesty's Stationery Office, 1966.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. The development of writing abilities (11-18). Urbana, Illinois: National Council of Teachers of English, 1975.

Brockriede, W.E. Dimensions of the concept of rhetoric. In K.K. Sereno & C.D. Mortenson (Eds.), Foundations of Communications Theory. New York: Harper and Row, 1970.

Brown, R. What we know now and how we could know more about writing ability in America. Journal of Basic Writing, 1978, 1, 1-6.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Christensen, F. A generative rhetoric of the sentence. College Composition and Communication, 1963, 24, 155-161.

Christensen, F. The problem of defining a mature style. English Journal, 1968, 57, 572-579.

Christensen, F. Sentence openers. College English, 1964, 24, 7-11.

Coffman, W.E. Essay examinations. In R.L. Thorndike (Ed.), Educational measurement (2nd. ed.). Washington, D.C.: American Council on Education, 1971. (a)

Coffman, W.E. On the reliability of ratings of essay examinations In English. Research in the Teaching of English, 1971, 5, 24-36. (b)

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.

Combs, W.E. Further effects of sentence-combining practice on writing ability. Research in the Teaching of English, 1976, 10, 137-149.

Cooper, C.R. Holistic evaluation of writing. In C.R. Cooper & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, Illinois: National Council of Teachers of English, 1977.

Cooper, C.R. & Odell, L. Introduction. In C.R. Cooper & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, Illinois: National Council of Teachers of English, 1977.

Crowhurst, M. & Piche, G. Audience and mode of discourse effects on syntactic complexity at two grade levels. Research in the Teaching of English, 1979, 13. 101-110.

CUNY Task Force on Writing. The CUNY writing skills assessment test: Student essays evaluated and annotated by the CUNY Task Force on Writing. Freshmen Skills Assessment Program, City University of New York, Spring 1979.

CUNY Task Force on Writing. Student information bulletin. Freshmen Skills Assessment Program, City University of New York, Fall 1978.

Diederich, P.B. Measuring growth in English. Urbana, Illinois: National Council of Teachers of English, 1974.

Diederich, P.B., French, S.W., & Carlton, S.T. Factors in the judgments of writing ability. Research Bulletin RB-61-15. Princeton, New Jersey: Educational Testing Service, 1961.

Dilworth, C.B., Jr., Reising, R.W., & Wolfe, D.T. Language structure and thought in written composition : Certain relationships. Research in the Teaching of English, 1978, 12, 97-106.

Dixon, E. Syntactic indexes and student writing performance. Elementary English, 1972, 49, 714-716.

Ebel, R.L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

Educational Testing Service. Guide to the use of the Descriptive Test of Language Skills. Princeton, New Jersey: College Entrance Examination Board, 1978.

Emig, J. The composing processes of twelfth graders. Urbana, Illinois: National Council of Teachers of English, 1971.

Emig, J. On teaching composition: Some hypotheses as definitions. Research in the Teaching of English, 1967, 1, 127-135.

Fagan, W.T., Cooper, C.R., & Jensen, J. Measures for research and evaluation in the English language arts. Urbana, Illinois: National Council of Teachers of English, 1975.

Follman, J.C., & Anderson, J.A. An investigation of the reliability of five procedures for grading English themes. Research in the Teaching of English, 1967, 1, 190-200.

Freedman, S.W. Influences on the evaluators of student writing. Unpublished doctoral dissertation, Stanford University, 1977.

Garcia, R.J. An investigation of relationships: Writing apprehension, syntactic performance, and writing quality. Unpublished doctoral dissertation, Arizona State University, 1977.

Gebhard, A.O. Writing quality and syntax: A transformational analysis of three prose samples. Research in the Teaching of English, 1978, 12, 211-232.

Godshalk, F., Swineford, F., & Coffman, W.E. The measurement of writing ability. New York: College Entrance Examination Board, 1966.

Golub, L.S. Syntactic density score (SDS) with some aids for tabulating. 1973. (ERIC Document Reproduction Service No. ED 091 741)

Golub, L.S. & Kidder, C. Syntactic density and the computer. Elementary English, 19[illegible] 51, 1128-1131.

Gronlund, N.E. Stating behavioral objectives for classroom objectives. New York: Macmillan, 1970.

Guilford, J.P. Psychometric methods (2nd. ed.). New York: McGraw-Hill, 1954.

Harris, W.H. Teacher response to student writing: A study of the response patterns of high school English teachers to determine the basis for teacher judgment of student writing. Research in the Teaching of English, 1977, 11, 175-185.

Hartnett, C.G. Measuring writing skills. Advanced Institutional Development Program Two Year College Consortium. 1978. (ERIC Document Reproduction Service No. ED 170 014)

Hillocks, G., Jr. Another review of The Development of Writing Abilities (11-18). Research in the Teaching of English, 1979, 13, 284-288.

Hooks, J.N. Tests on the English language. In A.H. Grommon, R. Braddock, J.N. Hook, W.A. Jenkins, W. Loban, & A.C. Purves (Eds.), Review of selected published tests in English. Urbana, Illinois: National Council of Teachers of English, 1976.

Hunt, K.W. Early blooming and late blooming syntactic structures. In C.R. Cooper & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, Illinois: National Council of Teachers of English, 1977.

Hunt, K.W. Grammatical structures written at three grade levels. Champaign, Illinois: National Council of Teachers of English, 1965.

Hunt, K.W. Syntactic maturity in schoolchildren and adults. Monographs of the society for research in child development. Chicago University: University of Chicago Press, 1970.

Jurgens, J.M., & Griffin, W.J. Relationships between overall quality and seven language features in compositions written in grades seven, nine, and eleven. Nashville, Tennessee: George Peabody College for Teachers, 1970. (ERIC Document Reproduction Service No. ED 046 932)

Kerlinger, F.N. Foundations of behavioral research (2nd. ed.). New York: Holt, Rinehart and Winston, 1973.

Kincaid, G.L. Some factors affecting variations in the quality of students' writing. In R. Braddock, R. Lloyd-Jones, & L. Schoer (Eds.), Research in written composition. Champaign, Illinois: National Council of Teachers of English, 1963.

Lindquist, E.F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton, Mifflin, 1953.

Lloyd-Jones, R. Primary trait scoring. In C.R. Cooper & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, Illinois: National Council of Teachers of English, 1977.

Loban, W.D. Language development: Kindergarten through grade twelve. Urbana, Illinois: National Council of Teachers of English, 1976.

Maimon, E.P., & Nodine, B.F. Measuring syntactic growth: Errors and expectations in sentence-combining practice with college freshmen. Research in the Teaching of English, 1978, 12, 233-244.

Marzano, R., & DiStephano, P. Five empirically based composition skills. 1978. (ERIC Document Reproduction Service No. ED 162 337)

McCleary, W.J. A note on reliability and validity problems in composition research. Research in the Teaching of English, 1979, 13, 274-278.

McColly, W. What does educational research say about the judging of writing ability? The Journal of Educational Research, 1970, 64, 148-155.

Mellon, J.C. Transformational sentence-combining. Champaign, Illinois: National Council of Teachers of English, 1969.

Meyers, L. Personal communication, Spring 1978.

Miller, B.D., & Ney, J.W. The effect of systematic oral exercises on the writing of fourth-grade students. Research in the Teaching of English, 1968, 2, 44-61.

Mitzel, H.E. Report on consulting visit concerning City University skills assessment program. Memorandum of the Vice-Chancellor for Academic Affairs, City University of New York, November 29, 1979.

Moffett, J. Teaching the universe of discourse. Boston: Houghton Mifflin, 1968.

Najjar, M.M. The relationship between syntactic development and performance by freshman students on selected competencies in English composition. Unpublished doctoral dissertation, Boston University, 1978.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. SPSS: Statistical package for the social sciences (2nd. ed.). New York: McGraw-Hill, 1975.

Nietzke, D.A. The influence of composition assignment upon grammatical structure. Unpublished doctoral dissertation, University of Illinois, 1971.

Nold, E.W., & Freedman, S.W. An analysis of readers' responses to essays. Research in the Teaching of English, 1977, 11, 164-174.

Odell, L. The classroom teacher as researcher. English Journal, 1976, 65, 106-111.

Odell, L. Measuring changes in the intellectual processes as one dimension of growth in writing. In C.R. Cooper & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, Illinois: National Council of Teachers of English, 1971.

Odell, L. Report on the CUNY writing skills assessment test. Memorandum of the Vice-Chancellor of Academic Affairs, City University of New York, November 29, 1979 (a).

Odell, L. Teachers of composition and needed research in discourse theory. College Composition and Communication, 1979, 30, 30-45 (b).

Odell, L., Cooper, C.R., & Courts, C. Discourse theory: Implications for research in composing. In C.R. Cooper & L. Odell (Eds.), Research on composing: Points of departure. Urbana, Illinois: National Council of Teachers of English, 1978.

O'Donnell, R.C. A critique of some indices of syntactic maturity. Research in the Teaching of English, 1976, 10, 31-38.

O'Hare, F. Sentence-combining: Improving student writing without formal grammar instruction. Urbana, Illinois: National Council of Teachers of English, 1973.

Page, E.B. The imminence of grading essays by computer. Phi Delta Kappan, 1966, 47, 238-243.

Perron, J.D. The impact of mode on written syntactic complexity: Part I--Third grade. Report No. 24. University of Georgia, 1976 (a). (ERIC Document Reproduction Service No. ED 126 531)

Perron, J.D. The impact of mode on written syntactic complexity: Part II--Fourth grade. Report No. 25. University of Georgia, 1976 (b). (ERIC Document Reproduction Service No. ED 126 511)

Perron, J.D. Written syntactic complexity and the modes of discourse. A paper presented at the annual meeting of the American Educational Research Association, April 1977. (ERIC Document Reproduction Service No. ED 139 009)

Pope, M. The syntax of fourth graders' narrative and explanatory speech. Research in the Teaching of English, 1974, 8, 219-227.

Rosen, H. An investigation of the effects of differentiated writing assignments on the performance in English composition of a selected group of 15/16 year old pupils. Unpublished doctoral dissertation, University of London, 1969.

Rothkopf, E., & Thurner, R. Effects of written instructional material on the statistical structure of test essays. Journal of Educational Psychology, 1970, 61, 83-89.

Sanders, N.M. Classroom questions: What kinds? New York: Harper and Row, 1966.

San Jose, C. Grammatical structures in four modes of writing at the fourth grade level. Unpublished doctoral dissertation, Syracuse University, 1972.

Seegers, J.C. Form of discourse and sentence structure. Elementary English Review 1933, 10, 51-54.

Shaughnessy, M.P. Basic writing. In G. Tate (Ed.), Teaching composition: Ten bibliographic essays. Fort Worth, Texas: Texas Christian University Press, 1976.

Shaughnessy, M.P. Errors and expectations. New York: Oxford University Press, 1977.

Simpson, G.F. Measures of writing ability of fourth, fifth, and sixth grade children. Unpublished doctoral dissertation, Kent State University, 1973.

Slotnick, H.B. & Rogers, W.T. Writing errors: Implications about student writers. Research in the Teaching of English, 1973, 7, 387-398.

Smith, C.T. The relationship between the type of questions, stimuli, and the oral language production of children. Research in the Teaching of English, 1977, 11, 111-116.

Stallard, C.K. An analysis of the writing behavior of good student writers. Research in the Teaching of English, 1974, 8, 206-218.

Stewart, M.F. Syntactic maturity from high school to university: A first look. Research in the Teaching of English, 1978, 12, 37-46.

Stewart, M.F. & Grobe, C.H. Syntactic maturity, mechanics of writing, and teachers' quality ratings. Research in the Teaching of English, 1979, 13, 207-215.

Veal, L.R. Syntactic measures and rated quality in the writing of young children. Studies in Language Education, Report No. 8. University of Georgia, 1974. (ERIC Document Service Reproduction No. ED 090 555)

Veal, L.R. & Tillman, M. Mode of discourse variation in the evaluation of children's writing. Research in the Teaching of English, 1971, 5, 37-45.

Watkins, F.C., Dillingham, W.B., & Martin, E.T. Practical English Handbook (4th ed.). Boston: Houghton Mifflin, 1974.

Whale, K.B., & Robinson, S. Modes of students' writing: A descriptive study. Research in the Teaching of English, 1978, 12, 349-355.

White, E.M. (Ed.) Comparison and contrast: The California State University and Colleges English Equivalency Examination. Long Beach, California: Office of the Chancellor, 1977.

White, E.M. Mass testing of individual writing: The California model. Journal of Basic Writing, 1978, 1, 18-38.

Williams, J.M. Re-evaluating evaluating. Journal of Basic Writing, 1978, 1, 7-17.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wolk, A. The relative importance of the free final modifier: A quantitative analysis. Research in the Teaching of English, 1970, 4, 59-68.

Zibrin, M. The 1979 audit of the writing assessment test. New York: City University of New York, 1979.

APPENDICES

APPENDIX A: Sample FSAP essay form. . . . . p. 76

Evaluation Scale. . . . . p. 77

Annotated student essays. . . . . p. 78

APPENDIX B: Essay questions used in study . . . p. 94

APPENDIX C: Inter-coder reliability data. . . . p. 98

APPENDIX A

THE CITY UNIVERSITY OF NEW YORK

WRITING SKILLS ASSESSMENT TEST

DIRECTIONS

You will have fifty minutes to plan and write the essay assigned below. (You may wish to use your fifty minutes in the following way: 10 minutes planning what you are going to write; 30 minutes writing; 10 minutes rereading and correcting what you have written.)

You should express your thoughts clearly and organize your ideas so that they will make sense to a reader. Correct grammar and sentence structure are important.

Write your essay on the lined pages of your booklet. You may use the inside of the front cover of the booklet for preliminary notes.

You must write your essay on one of the following assignments. Read each one carefully and then choose either A or B.

A. It always strikes me as a terrible shame to see young people spending so much of their time staring at television. If we could unplug all the TV sets in America, our children would grow up to be healthier, better educated, and more independent human beings.

Do you agree or disagree? Explain and illustrate your answer from your own experience, your observations of others, or your reading.

B. Older people bring to their work a lifetime of knowledge and experience. They should not be forced to retire, even if keeping them on the job cuts down on the opportunities for young people to find work.

Do you agree or disagree? Explain and illustrate your answer from your own experience, your observations of others, or your reading.

EVALUATION SCALE FOR WRITING ASSESSMENT TEST

6

The essay is competently organized and the ideas are expressed in appropriate language. A sense of pattern or development is present from beginning to end. The writer supports assertions with explanation or illustration.

Sentences reflect a command of syntax within the ordinary range of standard written English. Grammar, punctuation, and spelling are generally correct.

5 - 4

The writer introduces some point or idea and demonstrates an awareness that development or illustration is called for.

The essay presents a discernible pattern of organization, even if there are occasional digressions.

The essay demonstrates sufficient command of vocabulary to convey, without serious distortion or excessive simplification, the range of the writer's ideas.

Sentences reflect a sufficient command of syntax to ensure reasonable clarity of expression. The writer generally avoids both the monotony of rudimentary syntax and the incoherence created by tangled syntax

The writer demonstrates through punctuation an understanding of the boundaries of the sentence.

The writer spells the common words of the language with a reasonable degree of accuracy. Exceptions can be made for the so-called spelling "demons" which frequently trouble even an advanced writer.

The writer shows the ability to use regularly, but not necessarily faultlessly, the common forms of agreement and of grammatical inflection in standard written English.

3 - 2

An idea or point is suggested, but it is undeveloped or presented in a purely repetitious way.

The pattern of the essay is somewhat random and relationships between sentences and paragraphs are rarely signaled.

The essay is restricted to a very narrow range of language, so that the vocabulary chosen frequently does not serve the needs of the writer.

The syntax of the essay is not sufficiently stable to ensure reasonable clarity of expression. The syntax often is rudimentary or tangled.

The writer frequently commits errors of punctuation which obscure sentence boundaries.

The writer spells the common words of the language with only intermittent accuracy.

The essay reveals recurrent grammatical problems; if there are only occasional problems, this may be due to the extremely narrow range of syntactical choices the writer has used.

1

The essay suffers from general incoherence and has no discernible pattern of organization. It displays a high frequency of error in the regular features of standard written English. Lapses in punctuation, spelling, and grammar often frustrate the reader. OR, the essay is so brief that any reasonably accurate judgment of the writer's competence is impossible.

ESSAY 6

[1]Television is a modern form of entertainment which has become so popular that almost every American family owns at least one T.V. [2]Many people enjoy watching movies and variety shows on television, in the comfort of their own homes. [3]If we didn't own televisions, we'd be going out more at night, spending money at theaters and seeing plays. [4]Although these are enjoyable forms of entertainment, it is usually more convenient and economical to stay at home. [5]Television also offers a variety of entertainment while a single theater does not.

[6] Since television has such a great popularity, educators are striving to offer shows which provide education and not just entertainment. [7] Young people, especially, are very attached to watching television, since they have grown up with it. [8] For this reason, it is important that the daily practice of television viewing be put to good uses.

[9]Unfortunately, television is often misused. [10] Young people are allowed to see and hear things which they don't understand, and this can be harmful. [11] This is why it is most important that parents supervise what their children watch. [12]The businessmen behind television, as with every form of entertainment, offer what the public wants to see. [13]This is how they make their money. [14] Therefore, parents and educators should encourage their children to watch shows which provide some educational as well as entertaining value.

[15]Children often choose to watch cartoons which provide little or no educational value. [16]Instead, these shows are usually based on some sort of violence. [17]Since children are easily swayed, these shows can prove to be harmful on a child's outlook on life and on his relationship with others. [18]Children should be encouraged instead to watch shows like Sesame Street,

which offers educational entertainment. [19]This is an excellent program: viewers can learn how to build relationships, how different nationalities can live together in harmony, how to count and read on an elementary level. [20]All these things are taught by colorful puppets and happy people and funny cartoons. [21]From my personal experience in watching several children viewing Sesame Street, the show is one of the best children's shows available.

[22]Children use their creativity which evolves from what they see and hear around them. [23]Parents should put good television programs to use by helping children to relate to the programs. [24]Of course, other sources of entertainment and education, such as books and outdoor trips are also important.

[25]Like anything else in life, too much television may be harmful. [26]But treated in the right respect and with proper guidance, television can be the great invention that it was meant to be.

* * *

Comment on Essay 6

This essay is a 6 because it has a clear and sustained pattern of organization. The writer's use of connecting words and phrases holds the argument together. For example, she* makes a general statement (sentence 18), a supporting point (sentences 19 and 20), and then a personal observation (sentence 21), drawing clear relationships among these levels of generality. She also understands and mentions an argument other people might raise to challenge her point (sentences 24-26).

The writer's sentences are varied in length and structure. She subordinates ideas in different ways (sentences 3-6) and varies sentence length, using short sentences to good effect (sentence 13). In sentence 5, she accurately balances two parallel ideas through sentence structure.

The essay does, however, have limitations. There are occasional cliches (sentences 19 and 26), the style is sometimes stilted or repetitious (sentences 8, 14, and 23), and some sentences are unnecessarily wordy (sentence 11). Subject-verb agreement is occasionally weak (sentence 18). In addition, the logic of some sentences is unclear (sentence 12).

*In commenting on essays 2, 3A, 4A, and 6 we have referred to the writers as "she." In commenting om essays 1, 3B, 4B, and 5 we have referred to the writers as "he."

ESSAY 5

[1] I disagree with this passage. [2] The writer of the passage believes that the children of America are in many ways harmed by watching television and would be better off without it. [3] I agree that excessive television watching is not good for anyone. [4] Children who watch too much TV tend to become lazy and show a lack of imagination, and therefore they are harmed by it. [5] Television can also benefit children in many ways. [6] They seem to learn things much quicker and at a younger age by watching educational TV shows such as "Sesame Street" and "Electric Company," than children who don't watch these shows. [7] Shows like "Sesame Street" teach them things through repetition and they catch a child's attention easily through their use of bright colors, cartoons, music and puppets. [8] Children enjoy these shows and also learn from them.

[9] I believe that children should be guided by their parents as to what shows they can and cannot watch. [10] Children should watch shows that are suitable for them, that is, simple enough for them to understand but not so simple that they are bored by it. [11] They should not be allowed to watch shows that are excessively violent because these shows leave a bad impression on them. [12] I don't think that this pertains just to children, though, because adults are affected by these shows too.

[13] From my own observations, I have concluded that children between the ages of three and ten watch the most television. [14] These are very important years because the child's whole character is formed during these years. [15] If the child watches too much TV it can be harmful to him, especially if he is watching "junk" TV shows, that is shows with no educational value whatsoever. [16] Later on, it seems that young adults seem to drift away from TV and don't start watching it to an excess again until they hit about thirty. [17] By this

time, many of them have children of their own and are sitting down to enjoy it with them.

[18] In conclusion, I think that television watching has both advantages and disadvantages but in my opinion, the advantages far outweigh the disadvantages. [19]Children can really benefit from watching TV and I think the children of the generations will be smarter than we could have ever dreamed possible.

* * *

Comment on Essay 5

This essay is a 5 because the writer states a position and supports it with general statements (sentences 6 and 12). He then lets at least one of these general statements (sentence 6) organize supporting concrete detail. (sentences 6-8). Furthermore, in supporting his position the writer recognizes that the issue is complex, and he addresses himself, at least in passing, to some of these complexities (sentences 10 and 16-17).

Most of the writer's sentences are clear and are neither too colloquial nor stuffy. Some sentences demonstrate some skill in balancing ideas and in graceful phrasing (sentences 2, 4, and 10).

But the essay is a 5 and not a 6 mainly because the writer has not collected all his material regarding the effect of TV on children's learning (sentences 6-8, 15, 19) into a single logical paragraph. And although the writer raises interesting issues about watching TV when he discusses the relationship between parents and children, he has not drawn these scattered thoughts together (sentences 9-12, 16-17). As a result, paragraphs two and three drift from their announced topic. Also, sentence 19 is not clearly expressed.

ESSAY 4A

[1] Television has a harmful effect on young people. [2] If a child is watching television he cannot explore the outside world for himself. [3] The violence on these shows could make him believe there is no good only bad and television takes away the child's ability to imagine, think or play.

[4] If a child is in the habit of watching television he cannot use his sense of curiosity to learn about things around him. [5] The television does this for him, he does not initiate this feeling on his own. [6] It might be good for a child to see a flower, for example, on t.v. but nothing can show how it really is, to take it in your own hands and describe it for yourself. [7] This is what t.v. is taking away from today's children. [8] Their sense to learn and to explore on their own. [9] These young people are depending on t.v. for something which they have already, but just do not use.

[10] Most of what is shown on television today is violence. [11] Since young people are watching so much t.v. violence could be the only thing that appeals to them and they could carry this destruction to their own every day lives. [12] Children watching violence so much could get used to it, not care about real murders or wouldn't think a second time about picking a gun and killing somebody because his hero "Kojak" constantly does it on t.v. and he wants to be just like him. [13] Young children might not know the difference between the good guy or bad ones and sometimes on t.v. today the bad guys are the heros the roles have changed considerably.

[14] Television takes away the child's ability to imagine, play and think for himself. [15] If a child is constantly watch t.v. he doesn't have time to imagine or play or think for himself. [16] Playing is good for a young child because he learns how to share and understand others by the use of toys. [17] By thinking a child can determine his values not use his favorite t.v. characters' values

which might be invalid ones. [18] Imagining is taken away from the child because of t.v. advancement children can now go off to those far of lands that they used to imagine.

[19] Television has a distructive effect on a child. [20] He becomes addicted to it and does not go, outside to explore. [21] The child can take on invalid values seen on t.v. like violence. [22] T.V. can take away the ability for the child to use his own mind in everyday life.

* * *

Comment on Essay 4A

This essay is a 4 and not a 3 because it has a clear pattern of organization. The writer follows a simple five-paragraph form. She states her position and gives three reasons to support it in the first paragraph. She develops each reason in the next three paragraphs. And she concludes by restating the introduction. Each paragraph holds together pretty well. Specific illustrations help develop the argument in the middle paragraphs. The meaning of most sentences is clear, and the sentences are varied enough in length and structure to keep the reader interested. The writer's vocabulary seems adequate to her needs.

Despite these strengths, however, this essay is not a 5 because the writer does not connect ideas from paragraph to paragraph. This lack of transition weakens the overall logic of the essay. In addition, sentence beginnings and endings are sometimes not clear (sentences 5, 6, 8, 19). The writer tries to put two or three ideas into the same sentence without sufficient skill in handling sentence structure to allow her to do something meaningfully (sentences 13, 14, 19). Occasionally the writer uses language imprecisely (sentence 21) and is repetitious (sentences 15 and 16). She also misspells some common words.

ESSAY 4B

[1] I agree strongly with the passage, simply because it is very true. [2] It is said that Americans spend about five hours a day watching television. [3] Young people can be influenced easily by television. [4] They tend to believe what they see. [5] One example of this would be an incident that happened in Florida. [6] A young boy was watching a popular police show and saw one person kill another, the boy, probably feeling it would be alright, murdered a friend of his.

[7] The passage also states that it is not healthy watching too much T.V. [8] I agree with this also. [9] People tend to just sit back and watch others, in sporting events for example. [10] From this people can become very lazy, which I feel is not at all healthy.

[11] The passage says that "If we could unplug all the T.V. sets in America, we would be better off. [12] I cannot believe this because without the use of television we are losing many things, such as educational shows for children. [13] They help give the child a better understanding of the way the world in which he lives in operates. [14] We see a variety of Nature programs which help us understand how other thing live in the world.

[15] I do agree that television can corrupt the minds of young people. [16] They can learn many improper things that they do not need in life.

[17] I remember when I was younger that I would think that if it was on television and other people do things like murders, robberies, hijacking and other thing of that nature, I would feel thing like that happen all the time and that it was proper.

[18] I also agree strongly with idea of having a family hour. [19] In this time children and young people can watch worthwhile programs. [20] When it gets later into the evening, programs come on which are made for the older and more mature person.

[21] In conclusion I would like to say that television has many good and bad points. [22]Television has been around for a long time and it will be around for many more years to come. [23]Maybe in the future there will be "perfect television, where there will be programs to suit the needs of everyone in America.

* * *

Comment on Essay 4B

This essay's major strength, and what makes it a 4 rather than a 3, is that its organization is simple and clear, and its sentence construction is relatively skillful. The writer develops a pattern, maintaining it through simple connecting sentences at the beginning of each paragraph. Each paragraph takes up a separate aspect of the statement in question and offers concrete detail to support that aspect. The writer also uses varied sentence patterns and a vocabulary well chosen for the topic he is writing about.

The essay is not a 5, however, because as a whole its argument is inconsistent. Overall its ideas are randomly organized. The writer starts out "strongly" agreeing, but then, beginning in sentence 11, he disagrees with the passage given in the test question. In sentence 18, he refers to an issue that the question does not raise, and he concludes the essay with a statement that carefully avoids taking sides at all.

Also, although there is no pattern of major error, the writer sometimes does not show sentence endings (sentence 6). He misuses punctuation (sentence 17) as well as singular and plural forms (sentences 14 and 17).

ESSAY 3A

[1] I disagree with the statement that television has a harmful effect on young people. [2] If we unplug all the TV set, it may result in an opposite way rather than it said in the passage. [3] This result maybe that children become ignorant and blind to the modern society. [4] Another words, they will not know what is going on in this world by means of news cast on television.

[5] However, it's undeniable to say television also has a harmful effect to the children because of some series shown on TV that deal with actions and violences. [6] But besides those series, there are programs that are shown to provide to the young people the necessary knowledge and education. [7] These programs are often seen on channel thirteen which is considered an educational broadcastin company. [8] Many programs shown on which are excellent and interesting, and providing real and scientific knowledges to the viewers, such as NOVA, VISA and many others. [9] I have been watching these programs many times, although not becoming omniscient, I now understand more, so do the other people, the gradual advancement and improvement of this world.

[10] It is not necessary to have this kind of programs only on this public TV company. [11] Good programs can also be seen on those commercial TV broadcasting companies too. [12] There are certain hours that those channels will present some special series that are good to be watched. [13] For example, There are programs like "The Last of the Wild", "In Search of . . .", etc. shown on these channels.

[14] Again, it is not necessary to watch just these excellent series. [15] There are news broadcasting on most of the channels too, and which gives the viewers the events of the world. [16] A television is good for those who are illiterary but able to listen and for those who are lazy enough to pick up a copy of a newspaper.

[17] It is true that children in recent years have imitated the characters on the TV shows and done the things these characters did. [18] Sometimes crimes are resulted. [19] It is also true that children become lazier on their homework but spend their times on TV. [20] Besides these weakness that the TV provides, remember, there are good programs described above. [21] There is only one way to stop the belief of a harmful effect of television, which is to find a way to halt those violent shows on TV and concentrate on those more educational programs.

* * *

Comment on Essay 3A

This essay is a 3 and not a 2 because the writer makes her point, sticks to it, and offers some specific details to support her views. She is able to introduce a number of different aspects of her subject and keep them in a reasonably clear relation to one another. The writer establishes useful paragraph divisions (with the exception of the misplaced sentence 5) and, in her final paragraph, she points out opposing points of view (sentences 17-19).

This essay is not a 4, because the details the writer introduces are unexpanded and therefore unexplained (sentences 8, 13, and 15). As a result, although the writer has many potentially interesting ideas, they are not well-developed.

In addition, the writer's control of language and sentence structure is inconsistent. Her sentences are often tangled because of incorrect verb forms or lack of punctuation (sentences 8, 9, and 18). She confuses singular and plural forms (sentences 2, 3, 4, 8, 10, 20), and misuses or omits articles (sentences 3, 6, 16). Although the writer has a generally adequate vocabulary, she frequently puts words together incoherently and in that way confuses the reader. For example, "it may result in an opposite way" (sentence 2), and "there are news broadcasting" (sentence 15).

ESSAY 3B

[1] I agree that television does have an harmful effect on young people. [2] The young people will watch television then to do homework or prepare themselves for a test. [3] They will make up excuses for not doing the homework or prepare themselves for the test, the excuse might be one of the following: it doesn't matter if I fail the test I will still pass the class, it's only one homework assignment I will be missing, or the teacher shouldn't fail me because I missed one or failed one test. [4] This is how the young people feel, they would watch television before they would read a book. [5] I know they feel that way because all my friends are like that and I use to be like that. [6] I use to watch television then to read a book or do homework until I realized how important an education was. [7] Watching television wouldn't make me do better in school it makes me do worse watching television wouldn't get me a job.

[8] Television changes some peoples life styles they try to be like the person on the television. [9] After the program is over they might say I want to be like him or her. [10] They don't realize that is a person playing the role of a character. [11] If the character does something the young person might think that it is done like that in real life. [12] The young person could get hurt doing something that they might seen on television. [13] Then they wouldn't understand why it don't happen the same way for them. [14] Until someone explains to them that certain this on television only happens on television.

* * *

Comment on Essay 3B

This essay is a 3 and not a 2 primarily because the writer uses language with relative ease. He begins by stating his position clearly, and he offers two separate arguments to support that position. Some of the sentences are relatively clear (sentences 1, 2). Spelling throughout the essay is reasonably accurate.

The essay is not a 4, in part, however, because within each paragraph the writer's efforts to develop ideas tend to lead him off the track (sentence 3) or to offer other arguments which he does not express clearly or develop well (sentences 7 and 11-14).

The writer also has only limited skill with sentence structure and punctuation (sentences 3, 4, 5, 7, 8), and only a fair command of subject-verb agreement (sentence 13) and verb forms (sentence 12). Finally, his vocabulary does not entirely suit his needs. This limitation, together with poorly constructed sentences, frequently causes the writer to repeat words and to make inexact statements (sentence 4-12).

ESSAY 2

[1] Recently I have watched some of television's top rate shows. [2]The shows have no educational value at all. [3]The violence and sex that the networks put out is outragous. [4]Not only that, but it has an overall bad afect on grades in school. [5]The time these shows are air, interfers with time for study, term projects book reports, and homework.

[6]I do however; disagree that everyone should pull the plug on television. [7]Television does have its' positive side. [8]With the PBS (Channel 13) education shows (Nova, etc). [9]Television is an excellent means of communacation. [10]Television news has more information plus films so you have a feeling of what's happening and so you would be bore.

[11]A solution is to allow maybe one hours *of air time for network show an evening on weekdays and unlimited air time for weekends and holidays.

[12]The television could if used properly new tool of education if it is used wisely and by programming more educational series.

[13]Television is also a great advance. It's a good feeling to come home and watch a show to relax you after a long days work.

*one hour of prime time

* * *

Comment on Essay 2

This essay is a 2 and not a 1 because the writer shows some ability to organize her ideas in paragraphs one through four. She uses connecting words at the beginning of paragraphs to suggest that her argument is logical (sentences 6 and 11). In addition, the essay has some good sentences, at least in the first half, and the writer has tried to use a vocabulary appropriate to her topic.

The essay is not a 3, however, in part because it does not begin with a clear statement of the writer's position. Its development of ideas is weak because the logic of paragraphs one and two does not lead to the conclusion drawn in paragraph three. The third and fourth paragraphs are each only one sentence long, leaving their basic ideas undeveloped. The fourth paragraph is a better summary of the writer's position than the last paragraph, which goes off the subject and offers more arguments in favor of watching television.

The sentence structure, relatively strong until sentence 10, then becomes increasingly confused (sentences 11 and 12). In addition, the essay contains errors of punctuation which obscure sentence beginnings and endings (sentences 6, 8, 12), misspellings of common words (throughout), errors in subject-verb agreement (sentence 3), and incorrect pronoun reference (sentence 4).

ESSAY 1

[1] I agree, "is harmful effect on young people." [2] Most children is more effect than adult because young children watch on T.V. like one Sample violent is bad for children eyes, someday - they mind think is fun; I know one neighbor she had a son and once he watch T.V. on the program he saw Batman and Robin climbing the building so he try it, he climb the rope and the rope broke lose and fell on the floor bleeding top on his head - they rush him to the hospital he had about 10 or 15 stitches. [3] He never doing again he know only just fanstasy on T.V.

[4] T.V. cause more trouble like putting some sample on children mind; They should take out more violent, too many police series on t.v. and they should put good thing on Television program for the children, the world is too many violent already. [5] like cop and Robber - someday they children grow up - they do same thing on T.V. [6] I wish parent do something about this; parent should watch they children to be sure children don't allow watch violent or other thing on T.V. parent should spend more time with the children such as: give something to do than watch T.V.; like game or something keep away from T.V. [7] I read in the magazine they said, "Rate to high on violent on T.V. should put down less violent on T.V. many people is complain all the time; I believe they should put no violent and other thing on T.V. for young people, T.V. is harmful on people mind, and children.

* * *

Comment on Essay 1

This essay is not a 2 for several important reasons. First, the writer does not clearly indicate sentence beginnings and endings, so it is often hard to make sense of what he says. Second, although the essay keeps to its topic, it is confused. For example, the writer says "violent is bad for children eyes" (sentence 2), yet follows this with an example that doesn't support that statement. In the second paragraph he suggests, "they should put good thing on Television program for children" (sentence 4), but does not follow this with examples of what he has in mind. And he attempts to suggest what parents can do to help, but does not give details.

Third, the writer's word choice and sentence structure are weak. Words missing from sentences hide his meaning, and his vocabulary is too limited to say what he seems to want to say. Moreover, the essay suffers from many errors in standard written English: tense (lines 3-6), subject-verb agreement (sentences 2-4), and pluralizing (lines 7-11). The writer uses capital and lower case letters inconsistently (sentences 2 and 5). He uses punctuation poorly: scattered use of the dash, semi-colon, and colon (sentences 2, 5, and 6), omission of the apostrophe (sentences 2, 4, and 7) omission of quotation marks (sentence 7).

Despite its many problems, reflected in the rating of 1, this essay has some strengths. It is unified, since it deals exclusively with the point that there is too much violence on television. The writer tries to give examples to support his agreement with the test statement. The essay has a rudimentary paragraph division, a sense of an ending, and relatively accurate spelling.

APPENDIX B

High Cognitive/Low Experiential Demands

(1) In most American colleges, students must pass required courses in English, math, and science before they are allowed to take courses in their major area of study. Instead of making all students attend all of their required courses, colleges should offer more independent study programs in which students could complete some of their courses on their own, working at their own pace.

Do you agree or disagree with this statement? In an essay of about 300 words, explain and illustrate your answer in detail.

(2) The traditional grading system used by most colleges includes A, B, C, D, and F grades. This should be changed to a Pass/Fail system for all courses because the traditional grading system forces most students to compete for good grades rather than to strive for knowledge and learning.

Do you agree or disagree with this statement? In an essay of about 300 words, explain and illustrate your answer in detail.

High Cognitive/High Experiential Demands

(1) In most American colleges, students must pass required courses in English, math, and science before they are allowed to take courses in their major area of study. Instead of making all of you attend all your required courses, college should offer you more independent study programs in which you could complete some of these courses on your own, working at your own pace.

Do you agree or disagree with this statement? In an essay of about 300 words, explain and illustrate your answer in detail.

(2) The traditional grading system used by most colleges includes A, B, C, D, and F grades. This should be changed to a Pass/Fail system for all of your courses because the traditional grading system forces you to compete for good grades rather than to strive for your own knowledge and learning.

Do you agree or disagree with this statement? In an essay of about 300 words, explain and illustrate your answer in detail.

Low Cognitive/Low Experiential Demands

(1) In most American colleges, students must pass required courses in English, math, and science before they are allowed to take courses in their major area of study. Some colleges, however, offer students the option of taking independent study programs in which they can complete some of these required courses on their own, working at their own pace.

In an essay of about 300 words, discuss independent study as an alternative to attendance at required courses. In your essay, you might wish to choose one of the following strategies: contrast taking an independent study program to attending a required course, or explain some reasons for taking required courses, or suggest some changes in the rules concerning required courses and give reasons for these changes.

(2) The traditional grading system used by most colleges includes A, B, C, D, and F grades. Some colleges, however, use a Pass/Fail

system for all courses in order to move the focus of an education away from getting a certain grade to striving for knowledge and learning.

In an essay of about 300 words, discuss the traditional grading system (A-F). In your essay, you might wish to choose one of the following strategies: contrast the traditional grading system to a different type of grading system, or explain some reasons for keeping the traditional grading system, or suggest some changes in the traditional grading system and give reasons for these changes.

Low Cognitive/High Experiential Demands

(1) In most American colleges, students must pass required courses in English, math, and science before they are allowed to take courses in their major area of study. Some colleges, however, offer you the option of taking independent study programs in which you can complete some of these required courses on your own, working at your own pace.

In an essay of about 300 words, discuss independent study as an alternative to attendance at your required courses. In your essay, you might wish to choose one of the following strategies: contrast taking an independent study program to attending a required course, or explain some reasons for taking required courses, or suggest some changes in the rules concerning required courses and give reasons for these changes based on your experience.

(2) The traditional grading system used by most colleges includes A, B, C, D, and F grades. Some colleges, however, use a Pass/Fail

system for all courses in order to move the focus of your education away from getting a certain grade to striving for knowledge and learning in your classes.

In an essay of about 300 words, discuss the traditional grading system (A-F). In your essay, you might wish to choose one of the following strategies: contrast the traditional grading system to a different type that you have experienced, or explain some reasons for keeping the traditional grading system, or suggest some changes based on your experiences.

APPENDIX C

Inter-Coder Reliability Data

The following is based on Ebel (1951, p. 423):

Reliability of a group of coders =

$$\frac{\text{Mean Square of Essays - Mean Square of Error}}{\text{Mean Square of Essays}}$$

OWQ $r = \frac{77.33 - 9.93}{77.33} = .87$

TUS $r = \frac{109.75 - 1.17}{109.75} = .98$

WDS $r = \frac{31,529.39 - 99.85}{31,529.39} = .98$

SUBCLS $r = \frac{24.67 - 4.27}{24.67} = .83$

WDSFFM $r = \frac{461.06 - 12.64}{461.06} = .97$

FRAGS $r = \frac{6.17 - 0.20}{6.17} = .97$

ROS $r = \frac{2.39 - 0.04}{2.39} = .98$

MMS $r = \frac{2.72 - 0.18}{2.72} = .93$

SVAES $r = \frac{6.50 - 0.60}{6.50} = .91$

WFES $r = \frac{24.06 - 2.32}{24.06} = .90$

I's $r = \frac{6.17 - 0.33}{6.17} = .95$